

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 November 2007 (22.11.2007)

PCT

(10) International Publication Number
WO 2007/134128 A2

(51) International Patent Classification:
G06F 12/06 (2006.01)

(21) International Application Number:
PCT/US2007/068597

(22) International Filing Date: 9 May 2007 (09.05.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/382,670 10 May 2006 (10.05.2006) US

(71) Applicant (for all designated States except US): **IN-QUIRA** [US/US]; 851 Traeger Avenue, Suite 125, San Bruno, CA 94066 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **BIERNER, Gann, Alexander** [US/US]; 851 Traeger Avenue, Suite 125, San Bruno, CA 94066 (US).

(74) Agent: **FORD, Stephen, S.**; Marger, Johnson & McCollom, P.C., 210 S.W. Morrison Street, Suite 400, Portland, OR 97204 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

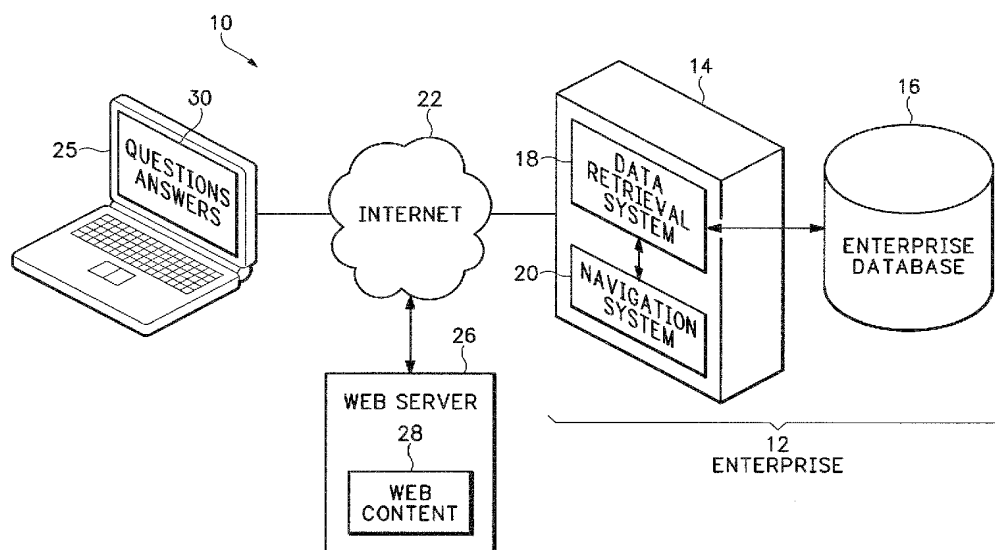
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: GUIDED NAVIGATION SYSTEM



(57) Abstract: A navigation system is used in combination with a data retrieval system to more effectively locate the correct answers or information in search content. The navigation system provides taxonomies that identify different information categories and sub-categories (facets). Classifiers automatically classify the search content by identifying the content corresponding with associated facets. The facets, in conjunction with the associated classifiers, provide robust classification and correlation between both structured and unstructured content. Rules can be used in conjunction with the facets and classifiers to automatically control different navigation operations. The navigation system can also leverage ontologies to dynamically create taxonomies and provides unique facet expansion and inlra-document classification operations.

WO 2007/134128 A2

GUIDED NAVIGATION SYSTEM

Cross Reference to Related Application

5 This application claims the benefit of U.S. Patent Application No. 11/382,670 filed May 10, 2006, and is herein incorporated by reference.

Background

 Current search engines receive a search request entered by a user and then identify and present the search results back to the search requester via a User Interface (UI). The search
10 results may be ranked in some order based on, for example, the presence of important words or concepts in the identified documents.

 Unfortunately, these search and display techniques often do not efficiently organize the information sought by the searcher. For example, the information or document sought by the user may be buried in one of the many different documents or data returned in the search results.
15 The user is then required to manually open and read dozens, if not hundreds, of the data items returned from the search results before identifying the correct, or most pertinent document.

 In yet another situation, the search requester may not know the correct search terms needed to direct or narrow a search to discover the desired document or information. Thus, the user is relegated to conducting multiple different time-consuming searches attempting to locate
20 the correct group of documents or information. The user has to then wade through all of the results from each search until a particular document or piece of information is discovered.

 The present invention addresses this and other problems associated with the prior art.

Summary of the Invention

A navigation system is used in combination with a data retrieval system to more effectively locate the correct answers or information in search content. The navigation system provides taxonomies that identify different information categories and sub-categories (facets). Classifiers automatically classify the search content by identifying the content corresponding with associated facets. The facets, in conjunction with the associated classifiers, provide robust classification and correlation between both structured and unstructured content. Rules can be used in conjunction with the facets and classifiers to automatically control different navigation operations. The navigation system can also leverage ontologies to dynamically create taxonomies and provides unique facet expansion and intra-document classification operations.

The foregoing and other objects, features and advantages of the invention will become more readily apparent from the following detailed description of a preferred embodiment of the invention which proceeds with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram of a network that uses a guided navigation system.

FIG. 2 is a diagram of a User Interface (UI) that is used with the navigation system shown in FIG. 1.

FIG. 3 is a diagram showing a partial example taxonomy.

FIG. 4 is a block diagram showing a partial taxonomy and classifiers used in the navigation system.

FIG. 5 shows a taxonomy builder used by a classifier to dynamically build taxonomies

from content.

FIG. 6 shows how the taxonomy builder builds a taxonomy from an ontology.

FIG. 7 shows how the navigation system can be used for intra-document classification.

FIGS. 8 and 9 show how the navigation system allows a user to expand a previous search
5 request.

FIG. 10 shows a parasitic classification scheme used in the navigation system.

FIG. 11 shows how custom classifier interfaces are used with the classifiers.

FIG. 12 shows another example of how an ontology is used in combination with search
content to automatically build a taxonomy.

10 FIGS. 13-16 show how rules are used to control different navigation operations.

FIG. 17 shows an alternative scheme for displaying both search answers and taxonomies
on the User Interface.

DETAILED DESCRIPTION

15 FIG. 1 shows a computer network system 10 that includes an enterprise 12 that has one or
more enterprise servers 14 and one or more enterprise databases 16 that contain content
associated with enterprise 12. For example, the enterprise 12 may be an on-line retailer that sells
books as well as other retail items. In this example, the enterprise database 16 may contain a
price list for all of the books available for purchase. In another example, the enterprise 12 may
20 be associated with a car manufacturer or car dealership and the enterprise database 16 could
include vehicle information. These are, of course, just two examples, and any type of business or
entity is represented by enterprise 12.

Other web servers 26 may operate outside of the enterprise 12 and may include associated

web files or other web content 28. Examples of content stored in enterprise database 16 and in file server 28 may include HTML web pages, PDF files, Word® documents, structured database information or any other type of electronic content that can contain essentially any type of information.

5 Some of the information may be stored in a structured format referred to generally as structured content. For example, data may be stored in the enterprise database 16 in a preconfigured format specified for enterprise 12. For example, a book or vehicle price list may be considered structured content. Alternatively, other information that is contained in enterprise database 16, or contained on other web servers 26, may be considered non-structured content.
10 This may include HTML web pages, text documents, or any other type of free flowing text or data that is not organized in a preconfigured data format known by the navigation system 20.

 A search query may be initiated from a terminal 25 through a User Interface (UI) 30. The terminal 25 in one example may be a Personal Computer (PC), laptop computer 24, wireless Personal Digital Assistant (PDA), cellular telephone, or any other wired or wireless device that
15 can access and display content over a packet switched network. In this example, the search query is initiated from the UI 30 and transported over the Internet 22 to the enterprise server 14. The enterprise server 14 operates a data retrieval system 18 that may search for both structured and/or non-structured content either in the enterprise database 16 or web server 26 pursuant to the query initiated in UI 30.

20 A novel navigation system 20 is then used to classify and navigate through the search results provided by the data retrieval system 18. The information as classified by the navigation system 20, along with other taxonomy information, is then displayed on the UI 30.

 The enterprise server 14 includes one or more processors that are configured to operate

the data retrieval system and navigation system 20. The operations performed by the navigation system could be provided by software computer instructions that are stored in a computer readable medium, such as memory on server 14. The instructions are then executed by a processor in server 14.

5 It should be understood that the examples presented below are used for illustrative purposes only and the scope of the invention is not limited to any of the specific examples described below.

Classification of Content

10 FIG. 2 shows the User Interface (UI) 30 that displays the output from the navigation system 20 shown in FIG. 1. The UI 30 includes a section where the a user can ask a question 36. In this example, the user inputs the text DRAGON in an attempt to locate a book related somehow to dragons. For example, the subject matter, title, author, etc., of a book may include the name or relate to dragons.

15 The data retrieval system 18 in FIG. 1 conducts a search in the enterprise database 16, and possibly in other external web servers 26, for content related to the DRAGON question 36. The results of the search are displayed on UI 30 as answers 38. In this example, the answers 38 include both structured content 40 and possibly unstructured content 42. As described above, the structure content 40 may include information that is preconfigured into predefined fields. For
20 example, each entry in a book list contained in the enterprise database 16 may contain a cover field 44A that shows the cover of the book, an author field 44B that identifies the book's author, a title field 44C that identifies the book title, and a price field 44D that identifies the book price. The content for each one of these fields for books that satisfy question 36 are displayed as

structured content answers 40. Each structured content answer 40 could also contain a buy button field 44E used for purchasing the book.

The answers 38 may also contain unstructured content 42 that may not have pre-defined data fields. For example, a first book review 42A may have been located somewhere in the enterprise database 16 or on a web server 27 and may describe a book related in some way to the question DRAGON. Similarly, another review 42B may be identified on the enterprise database 16 or on another web server 26 that is somehow also related to the DRAGON question 36. This content 42 is considered unstructured or “free-flowing” because the content is not presented in any particular preconfigured format. For example, the reference to a book related to DRAGON may be located anywhere in the book reviews 42A and 42B.

Of particular interest in the UI 30 is the display of a taxonomy 34 that identifies different classifications and sub-classifications for the content in answers 38. These classifications and sub-classifications are alternatively referred to as facets. In this example, the taxonomy 34 includes a first price taxonomy 34A that includes a first set of facets 35A that are associated with book prices, a second year taxonomy 34B that includes a set of facets 35B associated with book publication dates, and a third author taxonomy 34C that includes a set of facets 35C associated with book authors.

Different taxonomies 34 may be preconfigured in the navigation system 20. The taxonomies 34 associated with question 36, or with a particular enterprise, are selected by the navigation system, or the user, and displayed in UI 30. In some applications, some or all of the taxonomy is dynamically generated by the navigation system. This will be described in more detail below.

An associated number 46 may be displayed that indicates the number of content items 40

and 42 that correspond with a particular facet 35. For example, the search results (answers 38) for DRAGON may have produced a total of 74 structured content items 40 and/or unstructured content items 42 related to DRAGON books in a price range between \$0-\$5. Similarly, the navigation system 20 may have identified a combination of twenty-two structured content items 40 and/or unstructured content items 42 related to books published between 1980 and 1989. A process referred to as classification automatically classifies the content used to produce the answers 38 with different facets 35 in taxonomy 34. Classification is also described in more detail below.

When a user selects one of the facets 35 in taxonomy 34, the navigation system 20 automatically displays the structured and unstructured content 40 and 42, respectively, classified for the selected facet 35. For example, selecting the \$5-\$10 facet 35A causes the navigation system to display twenty-six answers 38 that are all associated somehow with dragon books within a price range of \$5-\$10.

Taxonomies

FIG. 3 shows some of the taxonomies 34 from FIG. 2 in more detail. A taxonomy is a hierarchical representation of content based on some feature (e.g., price, genre, publication year, etc). Content objects such as documents, sentences, database entries, etc are associated with items of the taxonomy (facets) using classifiers. For any given item in the taxonomy 34, all objects classified in that item (facet) 35 can be implicitly classified for all ancestor items in the taxonomy. For example, in the taxonomy below, every Fantasy book is also a Speculative book and also a Fiction book.

Taxonomy:

Genre

Fiction

Mystery

Romance

Speculative

Science Fiction

Fantasy

Non Fiction

Biography

History

Science

As mentioned above, a facet 35 is another name for an item in a taxonomy (a "taxonomy item"). The facets 35 can be referred to by a string constructed from the names of the facets along the path from the root of taxonomy to the facet. For example, each name may be separated by a period, ".", and all other periods may be preceded by a backslash. For example, the 1980-89 facet in FIG. 3 might be referred to as "Books.Year.1980-89."

Seamless Association of Structured and Unstructured Content

FIG. 4 shows how classifiers are used to provide seamless association of different structured and unstructured content with the same facets. Many systems allow navigation through structured data, such as e-commerce sites. However, the navigation system 20 provides classification of both structured content 40 and unstructured content 42, such as web content 28 (FIG.1).

In FIG. 4, the question 36 is entered through the UI 30 shown in FIG. 2. The question 36 in one embodiment is analyzed by the data retrieval system 18 that may use an Intelligent Matching Language (IML) 50. The operation of IML 50 for retrieving data is described in co-

pending patent application Serial No. 10/820,341, filed April 7, 2004, entitled: AN IMPROVED ONTOLOGY FOR USE WITH A SYSTEM, METHOD, AND COMPUTER READABLE MEDIUM FOR RETRIEVING INFORMATION AND RESPONSE TO A QUERY, which is herein incorporated by reference.

5 Briefly, the IML 50 allows content searching for documents that contain the same “concept” that is described in the question 36, for example. The navigation system 20 then provides navigation of the content provided by the IML 50 or can be used with any other data retrieval system that does not use IML 50.

10 As described above in FIG. 1, the data retrieval system 18 obtains content 52 either from databases 16 associated with the enterprise 12 (FIG. 1) or from other non-enterprise locations via Internet 22. The content 52 can be any type of electronic information that may relate to the question 36. For example, the content 52 may be a Microsoft® Word®, PDF, Powerpoint®, HTML, RTF, or DBASE document, etc., or any other type of document, text, or data item that contains information associated with question 36. These of course are only examples of data
15 formats and it should be understood that the content 52 can be any type of electronically captured data or text identified by data retrieval system 18.

As previously mentioned in FIG. 2, the content 52 may include both structured content 40 and unstructured content 42. The structured content 40 has predefined data fields 44. In the book example, the data fields 44 in structured content 40A may contain the book author, book
20 title, and book price. Other structured content 40B may also be contained in content 52 that has a different prefigured data structure. As also mentioned above, the unstructured content 42 may simply be free flowing text or some other text or data that is not presented in a predefined structured format.

The navigation system 30 uses classifiers 54 and 56 to seamlessly associate both the structured content 40 and the unstructured content 42 with the same associated facets 35. The taxonomy 34 by itself is just an association of classification items and has no direct connection to the content 52. The connection between content 52 and specific facets 35 is provided by the classifiers 54 and 56. These classifiers 54 and 56 include procedures that identify the pieces of content 52 that are associated with particular facets 35.

For example, the structured content classifier 54 may be associated with price facet 35A that identifies a book price range category of \$0-\$5. Accordingly, the structured classifier 54 may include procedures that access the price field 44D in structured content 40A and identifies any books with a price between \$0-\$5.

Of particular interest is the association of the structured classifier 54 and the unstructured classifier 56 with the same facet 35A. Unstructured classifier 56 includes procedures that parse through the unstructured content 42 for any documents that may indicate a book price in a range between \$0-\$5. For example, the unstructured content classifier 56 may parse through web pages to identify metadata that refers to a book having a price between \$0-\$5.

In one example, each piece of content 52 is assigned an associated Identifier (ID) 58 during the data retrieval process by data retrieval system 18. Upon identifying content 52 corresponding with the associated facet 35, the classifiers 54 and 56 then load the document ID 58 for the identified content into an associated facet list 59A. There are ID lists 59A-59C associated with each facet. In this example, the unstructured document 42A is identified by unstructured classifier 56 as associated with facet 35A and the associated ID 58A entered into list 59A. Similarly, the structured document 40B is identified by structured classifier 54 as

associated with facet 35A and the associated ID 58B entered into facet list 59A.

When the taxonomy 34 is displayed on UI 30 in FIG. 2, the number of content IDs for both structured and unstructured content that are possible results for the user's query are counted and displayed as document numbers 46 in FIG. 2. Thus, both structured and unstructured content 40 and 42, respectively, is seamlessly associated with the same facet 35 using classifiers 54 and 56.

It should be understood that any combination of different classifiers 54 and 56 can be associated with any combination of facets 35. In the example above, both a structured classifier 54 and an unstructured classifier 56 are each associated with a same facet 35A. However, there could be additional structured or unstructured classifiers associated with the same facet 35A. For example, facet 35A may use another, different structured classifier to classify other structured content 40B. Further, other types of unstructured classifiers may be associated with facet 35A to parse through different types of unstructured content 42.

There also may be one classifier associated with multiple different facets 35. For example, a classifier may be abstracted to identify any content 52 that indicates a book price. A taxonomy builder used by the classifier may then further refine classification of the identified content down to specific price ranges associated with the different price facets 35A, 35B, and 35C. This is described in more detail below.

In another embodiment, a same classifier may include procedures for identifying facet related content in both the structured and unstructured content 40 and 42, respectively. However there may be advantages to using separate structured and unstructured classifiers such as reduced software modification and increased software efficiency for different structured and unstructured data formats.

Some of the classifiers, such as unstructured classifier 56 may also use the IML 50 to more effectively identify content 52 associated with a particular facet 35. For example, the intelligent association of different concepts with a particular facet 35 allows the classifier 56 to identify content 42 that may not necessarily use the exact same words used in the facet. For instance, the IML 50 in unstructured classifier 56 may be used to identify unstructured content 42 that contains a book price but does not necessarily use the symbol "\$", or use the word "dollar". For example, the IML 50 may associate the symbol "\$" and the word "dollar" with the words "Euro", "bucks", "cost", "Yen", etc. The facet 35 may then simply apply a concept such as <dollar> or <price>. The IML 50 in classifier 56 which then automatically identifies any other documents that contain words associated with the <price> concept.

The classifiers in this example are associated with book information. But, of course, as previously mentioned above, the taxonomy, facets and associated classifiers are adaptable to any subject matter.

Taxonomy Builder

FIGS. 5 and 6 show how classifiers can also, though taxonomy builders, automatically create a taxonomy below an associated facet. Sometimes it is difficult or impossible to create a taxonomy completely by hand. For example, it may be desirable to classify a database based on a "category" and a "subcategory" field. One way to do this would be to create each facet by hand and add a classifier to select the appropriate rows out of the database. However, this may require hundreds of facets in such a taxonomy that, furthermore, might need to be changed at any notice when there is a database format change.

For this reason, some classifiers may not just identify the content associated with a

particular facet, but may also further subdivide the identified content 52 into subparts. This is performed by a taxonomy builder. Alternatively, a custom taxonomy builder can be built and used. The taxonomy builder takes the data from a classifier and uses it to construct a new taxonomy that is attached underneath an existing facet. Each time the content is classified, this taxonomy is reconstructed based on any new information.

For example, FIG. 5 shows one example where an existing taxonomy 60 may include price, year, and author facets 61A, 61B and 61C, respectively, as described above in FIG. 4. However, the price facet 61A may not have the sub-category facets 35A, 35B, and 35C that previously existed in FIG. 4. Alternatively, a taxonomy builder 62 is used by a classifier 61 associated with the price facet 61A. The classifier 61 searches content 52 for any documents that identify the price of books.

The procedures in the taxonomy builder 62 then automatically classify any identified content into different sub-categories such as \$0-\$5, \$5-\$10, \$10-\$15, etc. The taxonomy builder 62 then generates a taxonomy 66 under price facet 61A that includes new facets 68A-68D associated with different price ranges. The taxonomy builder 62 can then attach the newly created taxonomy 66 under the general price facet 61A previously created in taxonomy 60. The classifier 61 then classifies the content 52 for each newly built facet 68.

FIG. 6 shows another example of a taxonomy builder 74 used by a classifier 73. The taxonomy builder 74 uses an ontology 70 for automatically building a taxonomy 76. The ontology 70 is a hierarchical representation of a particular concept or subject matter. In one example, the ontology 70 is generated based on analysis of questions that may have been submitted to a particular enterprise site. Alternatively, the ontology 70 is a generic association of different meanings and words that may be associated with a particular concept. The use of the

ontology 70 is explained in further detail in the co-pending application Ser. No. 10/820,341, which has been incorporated by reference. The ontology 70 may have been previously created for use by the data retrieval system 18 for retrieving content 52 in response to question 36.

In this example, the same or a similar ontology 70 is also used for classifying the content 52 that was identified by the data retrieval system 18. The ontology 70 in this example is associated with the concept "vehicle model". The vehicle model concept may have multiple different categories, such as sedan, minivan, truck, etc. Under the sedan concept there may be multiple different sub-concepts that include Accord, Civic, and Taurus, etc.

Any concept or sub-concept can be referred to in IML 50. For example, the sub-concept Accord may be referred to as <Accord> using the IML 50. As described above, the IML 50 is used in the classifiers to automatically identify other words and phrases associated with the concept <Accord>.

The taxonomy builder 74 uses all, or a portion of the ontology 70, to automatically build a taxonomy 76. In this example, the taxonomy builder 74 extracts a section 72 of ontology 70 and inserts it into an associated sedan section 77 of vehicle model taxonomy 76. In this example, each sub-category in section 72 of ontology 70 is converted into an associated facet 78 in taxonomy 76.

The taxonomy builder 74 then automatically builds classifiers 79 for each of the automatically created facets 78A-78C in taxonomy 76. For example, an Accord unstructured content classifier 79A may be built by taxonomy builder 74 to classify unstructured data in content 52 associated with the concept <Accord> specified by facet 78A. Classifier 79A may be built by taxonomy builder 74 to use IML 50 to search for the concept <Accord> in the unstructured data in content 52. The taxonomy builder 74 may also automatically create an

Accord structured content classifier 79B to classify structured data in content 52 associated with the Accord facet 78A. Other classifiers may similarly be generated for the other automatically created facets 78B and 78C.

5 Intra-Document Classification

As described above, the navigation system 20 is not restricted to documents as the only level of search granularity. Content can also be classified in sentences, sections, or any other arbitrary document portion. This is especially useful for taxonomies or facets that are heavily represented in the content. For instance, an insurance company might have many documents that
10 each discuss different types of insurance. With document level classification, the user can restrict the query to, say, auto insurance but still get many other documents discussing other different types of unrelated insurance. For example, auto insurance may still appear
inconsequentially somewhere in many documents. With intra-document classification, a sentence or section specification increases the likelihood that all returned results are associated
15 with auto insurance.

To explain further, FIG. 7 shows a taxonomy 90 that is associated with vehicles. The taxonomy 90 may have different vehicle model categories that, in this example, include sedans, minivans and trucks. Any of the section or subsection facets can be specified at different intra-document granularities. For example, the facet 91 associated with the concept “minivan” may be
20 defined on a document, section, or sentence intra-document granularity. Of course other intra-document granularities can also be used and the three granularities described here are just examples.

The minivan document facet 91A is used by classifier 92 to identify any document in

content 96 that contains the concept <minivan> in a manner similar as discussed above.

However, the section facet 91B causes the classifier 92 to only identify document sections that contain the concept <minivan>. A document section could be defined for example as a paragraph, a grouping of text separated from other text by a line space, or any other delineation between groups of text. The sentence facet 91C only identifies sentences in content 96 containing the concept <minivan>. In the cases of sections and sentences, only answers that overlap these sections and sentences are shown to the user if that facet is selected.

To explain further, a user enters the text question “4-wheel drive”. The data retrieval system 18 retrieves content 96 that includes document A and document B. One or more intra-document classifiers 92 are associated with the minivan facet 91 and produce different results 98 depending on the intra-document classification used in facet 91.

For example, the minivan document classification facet 91A may be analyzed as the IML statement DOC(<minivan>) by classifier 92. Accordingly, classifier 92 identifies all documents in content 96 that include the concept <minivan>. In this example, both document A and document B would be identified in results 98A, since the concept minivan is contained in both 4-wheel drive documents.

The minivan sentence classification facet 91C is analyzed as the IML statement SENTENCE(<minivan>) by classifier 92. Classifier 92 identifies all sentences in content 96 that include the concept <minivan>. In this example, only document B includes both the concept <minivan> and the concept <4-wheel drive> in the same sentence and thus that answer, and not document A, is shown. Similar classification is used for the minivan section classification facet 91B where an IML statement SECTION(<minivan>) is used by classifier 92 to identify any sections within documents that contain the concept <minivan>.

Facet Expansion

In many query systems, it is difficult to back out of selections. The navigation system 20 (FIG. 1) allows a user to back out of a selection at any time to any level. The user can also expand the scope of a natural language query. For example, the user may enter the query "What books were published in 1992". After receiving search results, the user could then easily expand the scope of the query to include books published between 1990-1999. In another example, the navigation system 20 may receive a question for a particular car model. The navigation system 20 may identify content for different types of sedans. The user can then expand the scope of the query to other types of sedans after asking a first more specific question related to a particular car model.

FIG. 8 shows an example. A user may enter a question 104 via the UI 34, such as SUBARU OUTBACK WITH 4-WHEEL DRIVE. As discussed above, a data retrieval system 18 (FIG. 4) then searches one or more enterprise databases 16 and/or other sources on the Internet 22 for content associated with the question 104. The content is then displayed as answers 106.

In addition to the answers 106, the navigation system also displays a taxonomy 102 that is related to the question 104. In this example, the question is related to a particular vehicle model. Accordingly, the navigation system may present a model taxonomy 102 that includes the facets sedan and Outback, since Outback was requested by the question and sedan is its parent.

The navigation system provides the substantial advantage of allowing the user to then automatically expand the previous search by selecting one of the displayed facets. For example, the user may then select the facet 108 related to the broader category of sedans.

Referring to FIG. 9, the navigation system automatically rewrites the previous question to correspond to the selected facet 108. For example, the navigation system may rewrite the previous question SUBARU OUTBACK WITH 4-WHEEL DRIVE as a new question 110 that specifies SEDAN WITH 4-WHEEL DRIVE. The navigation system then automatically
5 resubmits the rewritten question 110 to the data retrieval system 18. The data retrieval system 18 may use the IML 50 to then conduct a new natural language query for content in the enterprise database 16 and possibly for other content located on the Internet 22. The new identified content 112 associated with sedans having 4-wheel drive may then be further classified by an associated sedan classifier 113 before being displayed as answers 114 on UI 34. Optionally, the UI 34 may
10 display the automatically rewritten question 110.

Parasitic Classification

The navigation system can take advantage of classifications in one data source to help classify content in other data sources. For example, an enterprise may have overlapping
15 structured and unstructured content. For instance, a book database may contain a structured price list and may also include reviews for some of the books identified in the price list. The navigation system 20 can not only classify both the structured and unstructured content from different databases but can use the structured book list database to further classify the unstructured content.

20 The navigation system may use a first structured classifier to identify books in the structured database within specific price ranges. The structured classifier does this by identifying book prices in a preconfigured price field. The classifier can then use other classifications in the structured database, such as the book title field, to identify books referred to in the non-structured

content that are within the same price range.

For example, the classifier can identify the titles in the structured database for all books within a particular price range. The identified book titles are then used by an unstructured classifier to identify similar book titles in the unstructured content. These book titles are then
5 classified as being within the same price range as the books identified in the structured content, even though the unstructured content may not identify or include a book price.

To explain further, FIG. 10 shows an information manager 122 that manages content, such as information related to different software releases. Different software products and related documents may be classified into different release versions 1.0, 2.0, 2.5, 3.0, etc. The
10 content in information manager 122 is considered at least partially structured since the data 124 is arranged into preconfigured version categories.

Also shown in FIG. 10 is unstructured content 120 that may contain different web documents that are also related to the same software product information contained in the information manager 122. This information 120 is considered unstructured since it is not
15 preconfigured into a predefined format or category groupings.

A structured classifier 126 generates a taxonomy 128 corresponding with the categories defined in structured content 124. For example, the taxonomy 128 may contain facets 129 that correspond to the different product release categories contained in structured content 124. The structured classifier 126 may then identify the data items in structured content 124 that are
20 associated with each release facet 129A-129D. For example, the classifier 126 may identify a document 125 contained in category “release 2.0” of content 124. The document 125 may describe a product referring to “turbo java”. Accordingly, the structured classifier 126 classifies the document 125 under the release 2.0 facet 129B.

Other classifiers, such as unstructured classifier 130, may then be associated with the facets 129 generated by structured classifier 126. The unstructured classifier 130 is created by the structured classifier 126 to search the unstructured content 120 for any information associated with the facets 129 generated from the structured content 124. For example, classifier 130 may search unstructured content 120 for any data items associated with the release 2.0 facet 129B. The unstructured classifier 130 may also use the IML 50 to identify any unstructured content 120 associated with concepts identified by the facets 129.

The classifier 130 can also be used to classify other content related to the facets 129. For example, a document in unstructured content 120 referring to “turbo java” may be classified under release 2.0 facet 129B due to the previous classification of the turbo java document 125 under the release 2.0 facet 129B by structured classifier 126. It should also be understood that the taxonomy classification could also be generated from unstructured content and then used to classify content in other unstructured or structured content.

Configurable Classifiers

FIG. 11 shows how classifiers can be customized for different database formats. For example, a taxonomy 140 and associated classifiers 142 may be developed for classifying a particular subject matter, such as cars. However, the content may be formatted differently for different databases 148 and 150 used by the same or different enterprises. For example, structured content 148 may associate different vehicle categories with different web pages. Other structured content 150 may format vehicle categories in a conventional column and row format.

Different custom classifier interfaces 144 and 146 convert the formats used in different content 148 and 150, respectively, into a common format that can then be used by the same

classifier 142. The classifier 142 can then classify both the content 148 and content 150 with associated facets in taxonomy 140. This prevents having to write different classifiers 142 for each different content format. Configurable classifiers are not restricted to structured data and may also be applied to unstructured data of any form.

5

Ontology/Result Based Classification

Referring to FIG. 12, the navigation system may automatically and dynamically generate taxonomies according to both returned results and a predetermined ontology. For example, a preconfigured ontology 160 may be associated with financial services. As described above, the ontology 160 could be manually generated according to empirical data, such as analysis of email questions received by a financial institution. In this example, the ontology 160 identifies different concepts associated with banking products. For example, these include, different types of checking and savings accounts, branch locations, and actions such as opening or closing an account.

10

A user may enter the question CHECKING ACCOUNTS via the UI 30 (FIG. 2). The data retrieval system 18 then retrieves content 162 from different databases using the IML 50. The content 162 may include different data items 163 that have some association with checking accounts. Some of the data items 163 may refer to a similar subject but use different terminology. For example, data item 163A refers to opening a checking account and data item 163F refers to creating a new checking account.

15

20

Simply using a taxonomy generator to identify common terms in the data items 163 and then creating an associated taxonomy may not yield the most effective content classification. For example, the taxonomy generator could create two separate facets, one for opening checking

accounts and another for creating new checking accounts. This could be confusing when viewed by the user.

Alternatively, the classifier 164 may not create a facet associated with new checking accounts because many of the data items use slightly different terminology when referring to the same general subject matter. Accordingly, the classifier 164 would not associate the content in any of the data items together and determine that there are not enough data items associated with any one of these terms to create a facet category.

To resolve this issue, the taxonomy builder 164 also refers to the ontology 160 when creating taxonomy 166. The taxonomy builder 164 analyzes the text of the results 162 as well as the concepts and other linguistic aspects in ontology 160. The concepts in ontology 160 may have different levels. For example, there may be customer concepts, industry concepts and language concepts. Certain concepts, such as customer concepts may be more important than industry concepts or language concepts. The taxonomy builder 164 compares the ontology 160 with the results 162 and builds a taxonomy 166 according to the comparison. For example, the taxonomy builder 164 may use known clustering algorithms that assign scores to the different concepts and words in ontology 160 corresponding with results 162. The taxonomy 166 is then generated based on this scoring.

It should be understood that taxonomy 166 is dynamically generated at run time based on the search results from a query. Other static taxonomies may also be used in the navigation system and created offline prior to operation.

To explain further, both the opening of a checking account referred to in data item 163A and creating a new checking account referred to in data item 163F may each be associated with a CHECKING ACCOUNT category 161A in ontology 160. If enough similar data items 163 are

associated with one of the same ontology category 161A, the classifier 164 may create a corresponding new account facet 168A in taxonomy 166.

Similarly, the taxonomy builder 164 may also generate a closing account facet 168B when enough data items 163 are associated with closing or shutting down a checking account and generate an interest rate facet 168C when enough data items 163 are associated with checking account interest rates. Note that content 162 has little or no data items associated with branch location category 161B in ontology 160. Accordingly, the taxonomy builder 164 does not create a corresponding facet in taxonomy 166. Thus, the ontology 160 and results 162 actually play off each other to create a better taxonomy classification.

Rules Based Facet Restriction

Rules can determine which facets are selected for classifying content. Users are not required to select facets in a taxonomy to allow particular facets to take effect. Rules allow the taxonomy to operate simply by the user asking a question. Rules can also be used to automatically select facets in the taxonomy according to different questions, even if the question does not exactly match an existing facet.

The rules include conditions that can refer to particular concepts in a question or particular types of users asking the question. Actions contained in the rules can then initiate a search for a particular concept or show a particular answer. The actions contained in the rules can also be used to manipulate facet restrictions.

FIG. 13 shows one example of a rule that translates questions into facet restrictions. In this example, a question 180 OUTBACK 4-WHEEL DRIVE is entered by a user. The data retrieval system 18 then identifies content 182 associated with question 180. The navigation

system also includes a taxonomy 170 that includes model facets associated with sedans, minivans and trucks as described above. Another section of the taxonomy 170 may also include pricing facets.

A rule 176 includes a condition 176A and an associated action 176B. In this example, the question 180 triggers the condition 176A in rule 176. For example, the condition 176A is premised on the concept <Outback> existing in the question 180. Note that a concept does not require the exact word "Outback" be used in the question 180 and only requires a word or phrase that is associated with the concept <Outback> as identified by an IML 50.

If the concept <Outback> exists in the question 180, the action 176B is triggered. In this example, the action 176B automatically restricts the search of content 182 to data items classified in a model.sedan.accord facet 171. Accordingly, only the answers 172 associated with the facet 171 are displayed.

FIG. 14 shows another example where a rule 184 includes a condition 184A that extracts a price #price=A - #price=A from a question 186. The associated action 184B initiates a price classification between \$12K and \$20K using the facets in price taxonomy 189. However, there is not a one-to-one correlation between the price range in condition 184A and the values for the prices facets 190 in price taxonomy 189. For example, the prices facets 190 specify \$5K price ranges \$10K-\$15K, \$15K-\$20K, and \$20K-\$25K.

The navigation system will then find the price facets 190 that cover the price range identified in question 186. For example, price facets 190A and 190B cover the entire specified price range of \$12K-\$20K. Accordingly, the price facets 190A and 190B are automatically selected in response to question 186 and the answers 188 classified from content 187 will only include vehicles having a price range of between \$10K-\$20K.

FIG. 15 shows another application for a navigation rule that associates particular actions with the type of user asking the question 192. For example, the rule 196 may have a business related condition 196A that is satisfied when the USER_ID associated with the user asking the question 192 is a gold card member. If the condition 196A is satisfied, the action 196B enables the display of answers 206B from both publicly available content and also answers 206A from subscription based morning star content. If the USER_ID is not associated with gold card membership in condition 196A, then only the public content is displayed in answers 206.

FIG. 16 shows yet another application of the navigation rules that display specific content according to facet selection. In this example, a rule 212 includes a condition 212A requiring selection of a particular Accord facet 220 in a vehicle model taxonomy 218. If a user satisfies condition 212A by selecting facet 220, the action 212B in rule 212 causes the navigation system to display an advertisement 224A for a particular type of LX Accord in answer 224.

Accordingly, a user may enter a question 210 and associated answers are displayed on the user interface 209. The user may then make a selection 222 for facet 220. The navigation system displays content 224B classified with Accord facet 220 and also displays the content in action 212B. In this case, the content in action 212B is an advertisement associated with facet 220.

Alternate Display Formats

FIG. 17 shows alternate ways that results and taxonomies can be displayed on a user interface. In this example, the same question DRAGON is entered by the user. Instead of the results being displayed in one section and the associated taxonomy displayed in a separate section, the answers and facets that the content is classified under are displayed together.

For example, the DRAGON question 252 is processed by the data retrieval system 18 described above. Some of the content from the data search is displayed as answers 254A and 254B. The answer 254A includes some of the same content that may be contained in the answers described above. However, the answer now also includes facets 256 and that have been
5 classified with the content.

For example, a classifier may have classified the content 256A and 256B with a \$0-\$5 price range facet 256C. Accordingly, the facet 256C is displayed along with content 256A and 256B. Similarly, the same, or another, classifier may have also classified the content 256A and 256B with a 1980-1989 publication date facet 256D. Accordingly, the facet 256D is displayed
10 along with the content 256A and 256B and facet 256C. Selecting the facet 256C or 256D in answer 254A operates in the same manner as the facets described above. Thus, selecting the year 1980-89 facet 256D would display answers for books published between the years 1980-89. Other answers, such as answer 254B, may display other facets that are associated with the corresponding content.

One of the advantages of this format is that the user is immediately notified that the book
15 identified in answer 254A is within a price range of \$0-\$5 dollars and published between the years of 1980-1989. This is displayed in answer 254A without ever clicking on a price range facet or clicking on a particular publication year facet.

In another variation, different taxonomy categories 260 may be displayed on the user
20 interface 250. The user can then select which taxonomies and search content is displayed as answers 254. An analog facet display 262 can also be used to select different sets of facets that correspond with the selected position on display 262.

Detailed Rule Description

The rules above are described below in more detail. Many of the descriptions below refers to the following taxonomies.

5 Taxonomy:

Genre
 Fiction
 Mystery
 10 Romance
 Speculative
 Fantasy
 Science Fiction
 Non Fiction
 15 Biography
 History
 Science

Price

20 \$0 - \$9.99
 \$0 - \$4.99
 \$5 - \$9.99
 \$9.99 - \$19.99
 \$10 - \$14.99
 25 \$15 - \$19.99
 \$20 - \$29.99
 \$20 - \$24.99
 \$25 - \$29.99

30 A facet can be automatically selected in a rule using an "Add Facet Restriction" item in an "Advanced" menu. This will create an action tab with two mechanisms for selecting facets. First, there is a text field in which the full name of the facet can be typed in quotes (e.g., "Genre.Fiction"). The appropriate facet within the tree shown below the text field can also be selected. Note that this tree may not be complete until the data is classified.

Multiple facets can be selected using rules. As with selecting one facet, multiple facets can be selected within a "Add Facet Restriction" action in either the text field or the tree. Within the text box, an IML expression can be written where the primary difference is that quotes text is interpreted as facet references rather than tokens.

Two facets, a and b, can be selected with the expression: "a" OR "b". If only results that were classified by both facets is desired, the following expression can be used: "a" OVERLAP "b". Note that "a" IS "b" will also work but only for exact matches. For example, all documents will be correctly returned that were classified as a and b. However, no match will occur for a document classified as a with a section classified as b. The OVERLAP operator will cause the section to be matched.

Selecting facets in the tree will produce an expression which is the intersection of the taxonomies and a union of the facets within the taxonomies. For instance, selecting Mystery, Fantasy, and \$20 - \$29.99 in the genre taxonomy, will produce the expression:

("Genre.Fiction.Mystery" OR "Genre.Fiction.Speculative.Fantasy") OVERLAP "Price.\$20-\$29.99".

Facets may be selected in different rules. The navigation system attempts to combine them in a reasonable way in the same order that the rules were analyzed. As long as the facet expressions are in the form of an intersection of unions and the unions are isolated to one taxonomy each, then facet expressions are combined to maintain that form. If a different type of expression is found, from that point on, expressions are combined using OVERLAP.

For example, consider the case where three rules fire with the following expression.
"a.b" OVERLAP ("x.y" OR "x.z")

"a.c" OVERLAP "1.2"

"a.b" OR "x.z"

The first two expressions conform to the standard expression style, so they combine into:
("a.b" OR "a.c") OVERLAP ("x.y" OR "x.z") OVERLAP "1.2". However, the third expression
5 does not conform (it contains a union from two different taxonomies) and thus is merely
appended with the OVERLAP operator: ("a.b" OR "a.c") OVERLAP ("x.y" OR "x.z")
OVERLAP "1.2" OVERLAP ("a.b" OR "x.z"). Note that facets selected in a search rule only
affect that search. Facets selected in other rules are combined and affect all subsequent searches.

A facet can be restricted for a particular search. Adding such a restriction within a search
10 rule may apply just to that search and may, by default, be combined with the previous
restrictions. An option can also be selected to not combine with previous restrictions. This is
particularly useful for doing collection based facets which are not to be applied to structured
searches.

A previous facet restriction can be over-ridden. A search rule may include the Facet
15 Restriction action and also a checkbox which, when checked, causes the current restriction to
replace previous restrictions.

Facet references can contain configurable wildcards if the full path of the facet is
unknown. For example, it may not be exactly known where Fantasy is located in the genre
taxonomy. Fantasy can be referred to using Genre.*.Fantasy. This returns the first facet in the
20 genre taxonomy that has the name Fantasy using a standard breadth first search. In other words,
the first, shallowest possible match, is identified. If there is no match, the facet will be removed
from the expression.

The id for a facet may not be known. This is sometimes the case with numeric or range

based facets where the facet reference cannot exactly be expressed. Consider the taxonomy of prices. A rule can match prices and create a facet restriction such as `Price.*.{price}`.

Unfortunately, individual prices may not exist in the taxonomy preventing a match. To find the most specific facet whose range contains the price, the following restriction can be used:

5 `Price.<*. {price}`. Note that this extension requires min/max values and a comparator to be specified for the facet. Some classifiers may do this automatically.

An approximate match may be generated for a facet. Sometimes a user may ask something like "Show me books around \$15". If the facet reference `Price.<*. {price}` is used, the facet \$15-\$19.99 may be returned back which is probably not the desired result. A similar
10 case may exist when a user asks "What books were published in 1998?". The preferred response may not be 1998, but 1990-1999. This can be accomplished with the `~*` wildcard.

For these two examples, that would be `Price.~*.$15` and `Year.~*.1998`.

To be precise, this expression looks for a containing facet first. If the requested value is completely contained in the facet, then that facet is matched (e.g., both `Price.<*.$12` and
15 `Price.~*.$12` will match \$10-\$14.99). Otherwise, the parent is matched. So, `Price.~*.$15` will match \$10-\$19.99. Note that this extension has min/max values and a comparator to be specified for the facet. Some classifiers may do this automatically.

A range of facets can be selected. For example, a user may ask "What books cost between \$0 and \$22". This range can be requested in a facet restriction expression by using an `*`
20 IML operator which is overloaded for this purpose. The following expression would be used: `"Price.<*.$0" * "Price.<*.$22"`. This will cause the facets \$0-\$9.99, \$10-\$19.99, and \$20-\$14.99 to be selected. Note that this extension has min/max values and a comparator to be specified for the facet. Some classifiers may also do this automatically.

The user may also request a value that doesn't exist. In many cases, the best thing to do is nothing, and the facet restriction will be ignored. But sometimes, especially when the user has asked for a range, some action is required. For example, the user may have asked for books starting at -10 dollars. Alternatively, the user may be asking for books published starting at 1945
5 when there is only data starting at 1960.

The following wildcards can be used. The "<*" wildcard is equivalent to <◇* but if no answer is returned, finds the closest match that is bigger than the requested value. For example, Price.<*. {price} matches \$0-\$4.99 if the {price} evaluates to \$-10, \$-3.19, \$0, \$2.37, etc. The ">*" operator is equivalent to >◇* but if no answer is returned, finds the closest match that is
10 smaller than the requested value.

The user can be restricted from seeing particular facets. A facet restriction can be set using ISNT. For example, the user should not see any non-fiction books, the following expression can be used: "Genre" ISNT "Genre.Non Fiction"

A user can be restricted to see only particular facets. For example, to restrict a user to
15 only fiction books, rather than using "Genre.Fiction", the following operator can be used "RESTRICTED(Genre.Fiction)".

A facet may be selected in the rules and the user may then navigate to a more general facet, the navigation system may still search for the original facet. To solve this problem, the navigation system can "fix" the question by replacing the old facet referring text with the new.
20 The user still sees the old question, but as far as the system is concerned, the question has been changed. A variable may be set on the facet restriction that corresponds to the part of the question that needs to be replaced.

For example, there may be a rule with the following language condition ("science

fiction")=scifi. The facet restriction would then be "Genre.*.Science Fiction"=scifi. Now the navigation system can tell what part of the question was about science fiction and replace it with the appropriate information.

The system described above can use dedicated processor systems, micro controllers,
5 programmable logic devices, or microprocessors that perform some or all of the operations. Some of the operations described above may be implemented in software and other operations may be implemented in hardware.

For the sake of convenience, the operations are described as various interconnected functional blocks or distinct software modules. This is not necessary, however, and there may be
10 cases where these functional blocks or modules are equivalently aggregated into a single logic device, program or operation with unclear boundaries. In any event, the functional blocks and software modules or features of the flexible interface can be implemented by themselves, or in combination with other operations in either hardware or software.

Having described and illustrated the principles of the invention in a preferred
15 embodiment thereof, it should be apparent that the invention may be modified in arrangement and detail without departing from such principles. I claim all modifications and variation coming within the spirit and scope of the following claims.

Claims

1. A method for classifying content, comprising:

receiving both structured and unstructured content;

5 providing a taxonomy having facets associated with different subject categories;

using a structured classifier to classify the structured content with one of the facets; and

using a unstructured classifier to classify the unstructured content with the facet.

2. The method according to claim 1 wherein the structured classifier searches predetermined

10 data fields or predefined locations in the structured content for information associated with the facet.

3. The method according to claim 1 wherein the unstructured classifier searches for

information associated with the facet that is not located in predefined data fields or predefined

15 locations in the unstructured content.

4. The method according to claim 3 including using an Intelligent Matching Language

(IML) in the unstructured classifier to search data items in the unstructured content

corresponding with a concept identified by the facet.

20 5. The method according to claim 1 including:

assigning identifiers to data items in both the structured and unstructured content;

using the structured classifier to generate a list of the identifiers for the data items in the

structured content associated with the facet;

using the unstructured facet classifier to add the identifiers to the same list for the data items in the unstructured content associated with the facet; and

displaying the number of data items in both the structured and unstructured content associated with the same facet by counting the number of identifiers in the list that are possible answers to a query.

6. The method according to claim 1 including:

identifying data items in either the structured or unstructured content associated with the facet;

identifying different sub-categories for the identified data items;

building a new taxonomy by listing the sub-categories underneath the facet; and

classifying the data items according to the new taxonomy.

7. The method according to claim 1 including:

associating an ontology with the structured and unstructured content that includes at least one concept having different categories and sub-categories;

inserting a portion of the ontology into the taxonomy; and

using the categories and sub-categories for the extracted portion of the ontology as facets for classifying the structured and unstructured content.

8. The method according to claim 7 including:

building new classifiers for the facets created from the ontology; and

using the new classifiers to classify the content with the facets created from the ontology.

9. The method according to claim 7 including:

conducting a query that returns search results;

5 comparing concept categories in the ontology with data items contained in the search results;

identifying the concepts categories associated with the data items from the search results;

and

dynamically creating facets in the taxonomy according to the concept categories in the

10 ontology corresponding with the search results.

10. A computing system, comprising:

a processor configured to receive both structured and unstructured content and classify both the structured and unstructured content into categories identified in an associated taxonomy.

15 11. The computing system according to claim 10 wherein the processor is further configured to independently operate both structured and unstructured classifiers for the same taxonomy categories, the structured classifiers classifying the structured content for associated taxonomy categories and the unstructured classifiers classifying the unstructured content for associated
20 taxonomy categories.

12. The computing system according to claim 10 wherein the processor is configured to identify at least some of the taxonomy categories as concepts and then operate an Intelligent

Matching Language (IML) that identifies the taxonomy category concepts with data items in the structured and/or unstructured content.

13. The computing system according to claim 10 including memory that contains lists
5 associated with the taxonomy categories, the processor configured to classify data items in both the structured and unstructured content with the taxonomy categories and then store identifiers for the classified items in the lists associated with the taxonomy categories.

14. The computing system according to claim 13 wherein the processor tracks a number of
10 identifiers contained in the lists associated with the taxonomy categories and displays the number on a User Interface (UI) along side the associated taxonomy categories.

15. The computing system according to claim 10 wherein the processor is configured to
15 automatically generate sub-categories for one or more of the taxonomy categories and then classify both the structured and unstructured content for both the taxonomy categories and taxonomy sub-categories.

16. The computing system according to claim 15 wherein the processor is further configured
20 to automatically create classifiers for the generated taxonomy sub-categories that then automatically identify data items in the structured and unstructured content associated with the taxonomy sub-categories.

17. The computing system according to claim 10 wherein the processor is configured to provide an ontology that includes different ontology categories and ontology sub-categories, the processor inserting a portion of the ontology into the taxonomy and classifying the structured and unstructured content using the categories and sub-categories extracted from the ontology.

5

18. The method according to claim 17 wherein the processor is configured to compare concepts identified in the ontology with data items in the structured and/or unstructured content and identify the concepts in the ontology that are associated with the data items, the processor then configured to automatically create the taxonomy according to the identified concepts in the ontology.

10

19. A method for building taxonomies in a document navigation system, comprising:
identifying one or more taxonomy categories for classifying content received back from a query;

15 automatically building taxonomy sub-categories from a portion of an existing ontology to provide sub-classifications of the identified taxonomy categories; and

using the taxonomy categories and built taxonomy sub-categories to automatically classify the content received back from the query.

20 20. The method according to claim 19 including identifying the one or more taxonomy categories from an existing taxonomy and then dynamically generating the taxonomy sub-categories at run time from data items received back from a query and initially classified using the existing taxonomy categories.

21. The method according to claim 19 including building classifiers that automatically classify the content for the built taxonomy categories and/or sub-categories.

5 22. The method according to claim 21 including:

identifying data items in the content associated with the ontology;

analyzing the different types of concepts and/or words in the ontology that are associated with the identified data items; and

10 dynamically creating a taxonomy according to the different types of concepts and/or other types of non-concept words in the ontology that are associated with the data items.

23. A computer readable medium containing instructions that when executed, comprise:

receiving both structured and unstructured content;

15 providing a taxonomy having facets associated with different subject categories;

using a structured classifier to classify the structured content with one of the facets; and

using a unstructured classifier to classify the unstructured content with the facet.

24. The computer readable medium according to claim 23 wherein the structured classifier

20 searches predetermined data fields or predefined locations in the structured content for information associated with the facet.

25. The computer readable medium according to claim 23 wherein the unstructured classifier

searches for information associated with the facet that is not located in predefined data fields or predefined locations in the unstructured content.

26. The computer readable medium according to claim 23 including instructions that use an Intelligent Matching Language (IML) in the unstructured classifier to search data items in the unstructured content corresponding with a concept identified by the facet.

27. The computer readable medium according to claim 23 including instructions that when executed:

identify data items in either the structured or unstructured content associated with the facet;

identify different sub-categories for the identified data items;

build a new taxonomy by listing the sub-categories underneath the facet; and

classify the data items according to the new taxonomy.

28. The computer readable medium according to claim 23 including instructions that when executed:

associate an ontology with the structured and unstructured content that includes at least one concept having different categories and sub-categories;

insert a portion of the ontology into the taxonomy; and

use the categories and sub-categories for the extracted portion of the ontology as facets for classifying the structured and unstructured content.

29. The computer readable medium according to claim 28 including instructions that when executed:

build new classifiers for the facets created from the ontology; and
use the new classifiers to classify the content with the facets created from the ontology.

5

30. The computer readable medium according to claim 29 including instructions that when executed:

conduct a query that returns search results;
compare concept categories in the ontology with data items contained in the search

10 results;

identify the concepts categories associated with the data items from the search results;
and

dynamically creating facets in the taxonomy according to the concept categories in the
ontology corresponding with the search results.

15

1/17

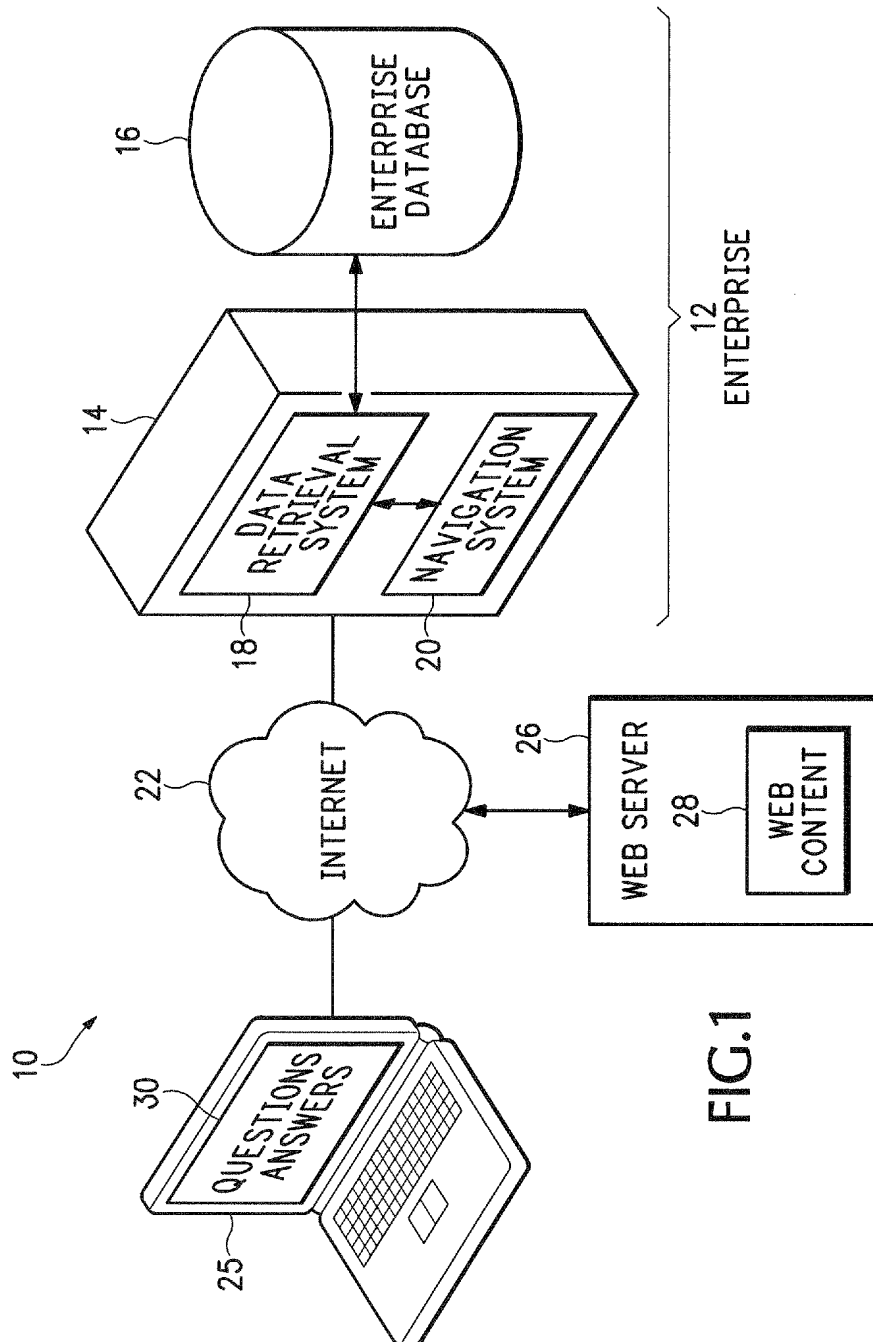


FIG.1

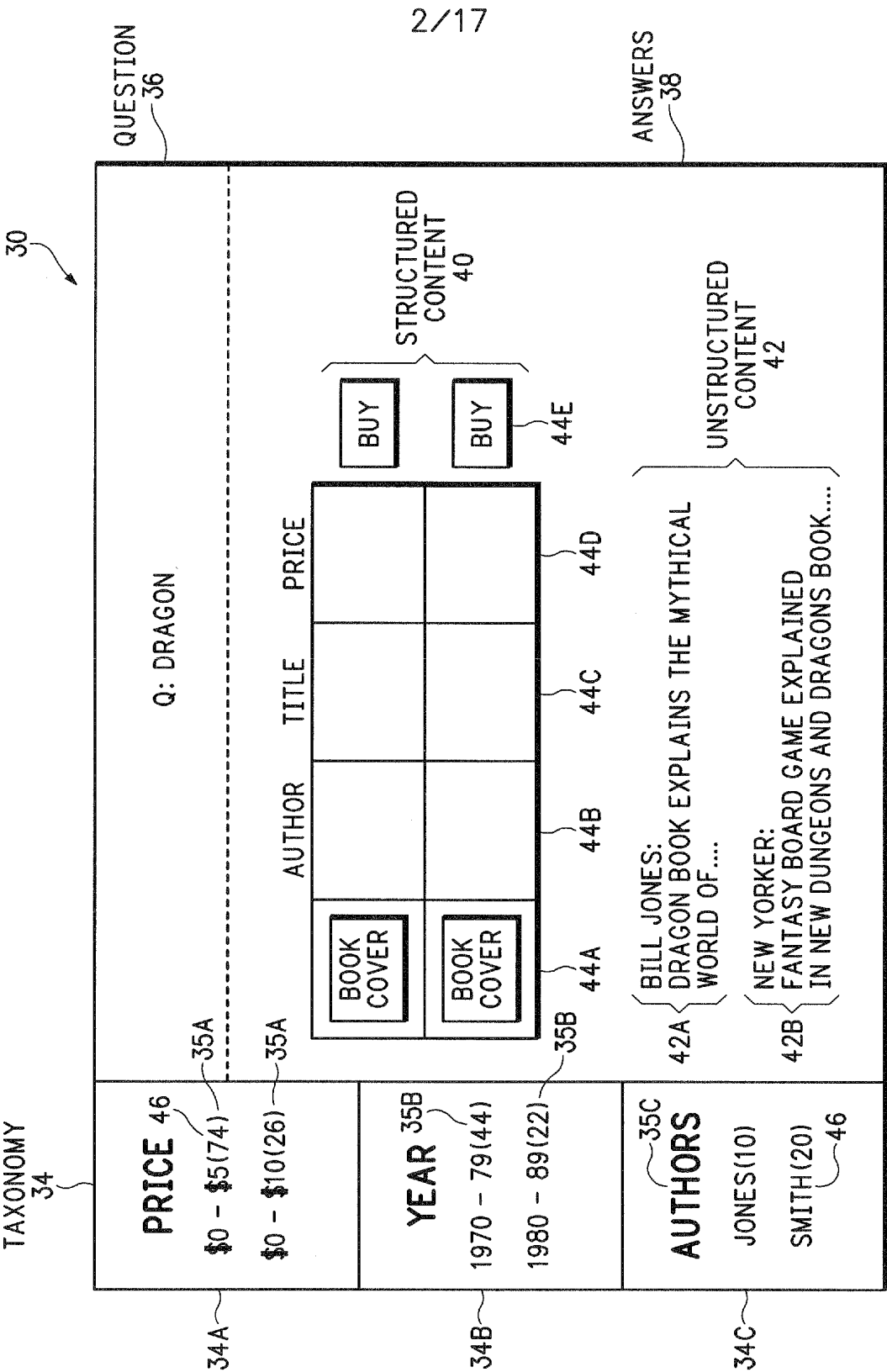


FIG.2

3/17

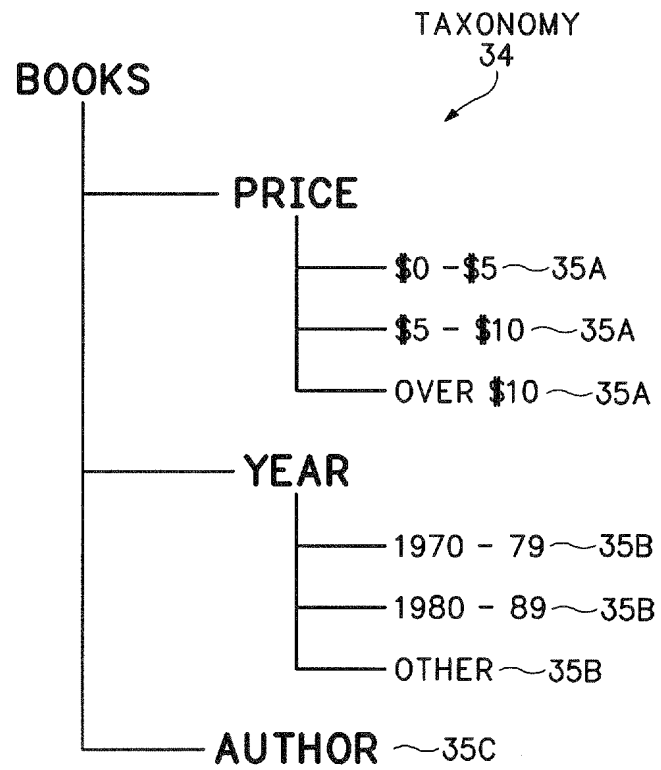
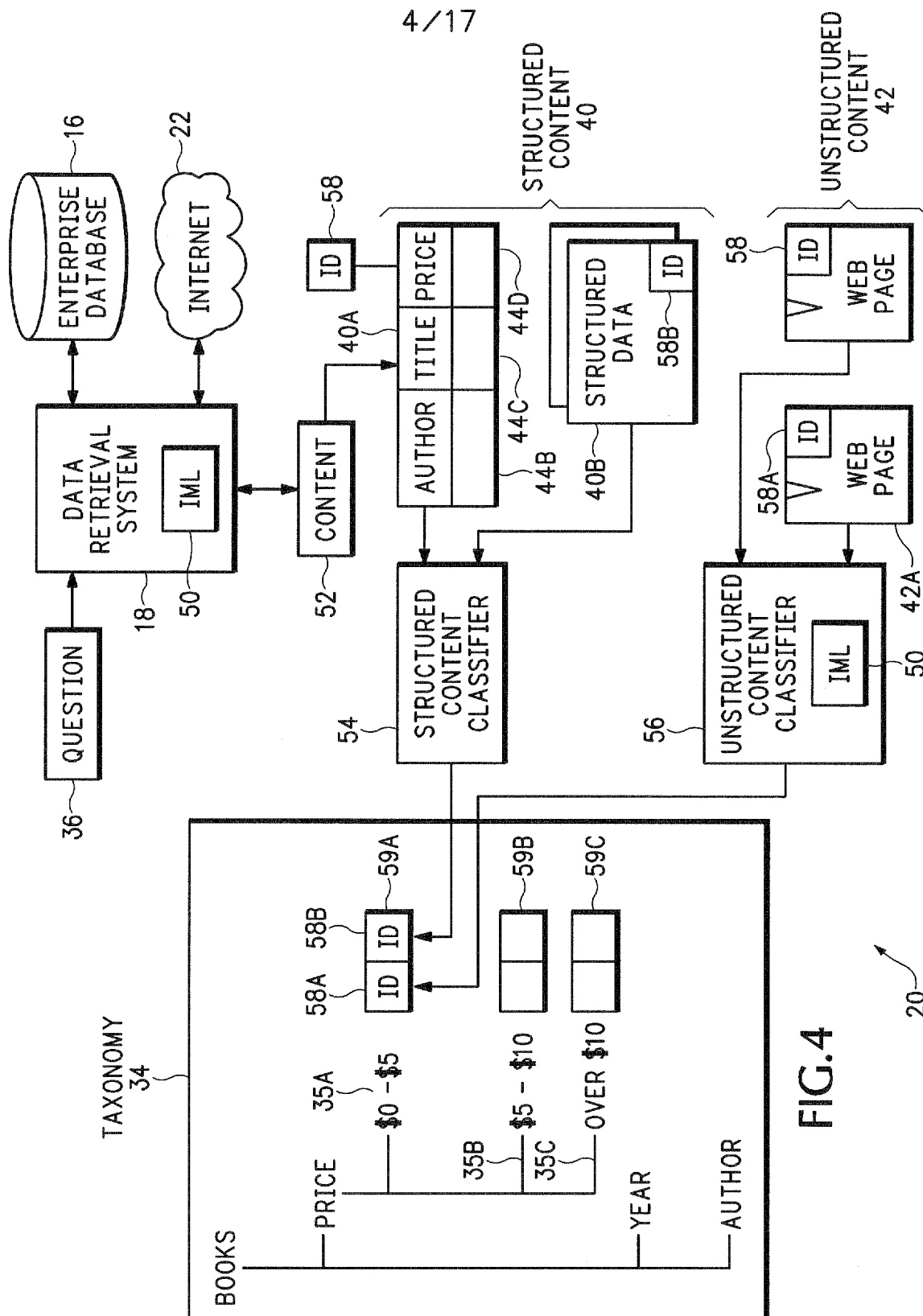


FIG.3



5/17

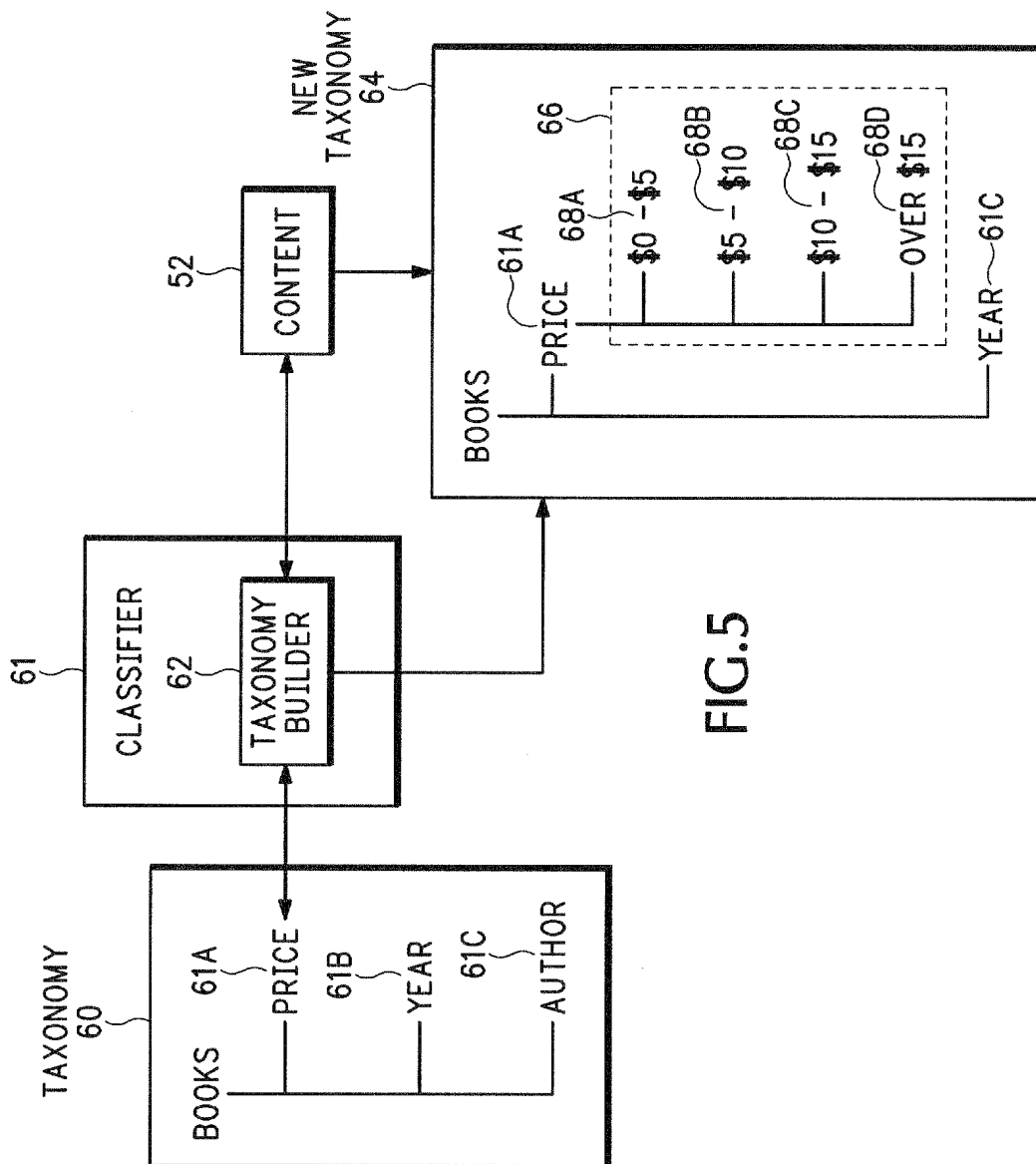


FIG.5

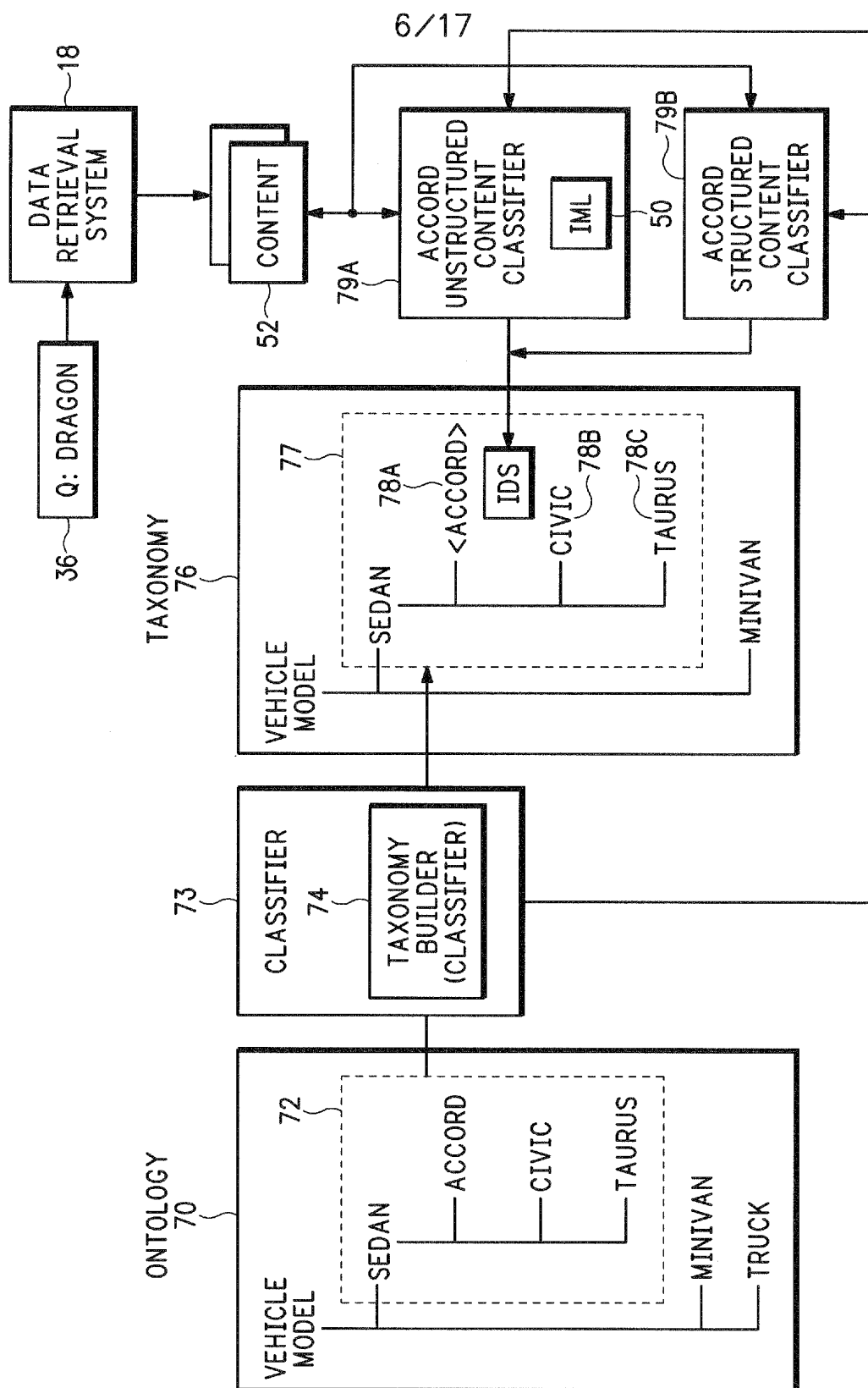


FIG.6

7/17

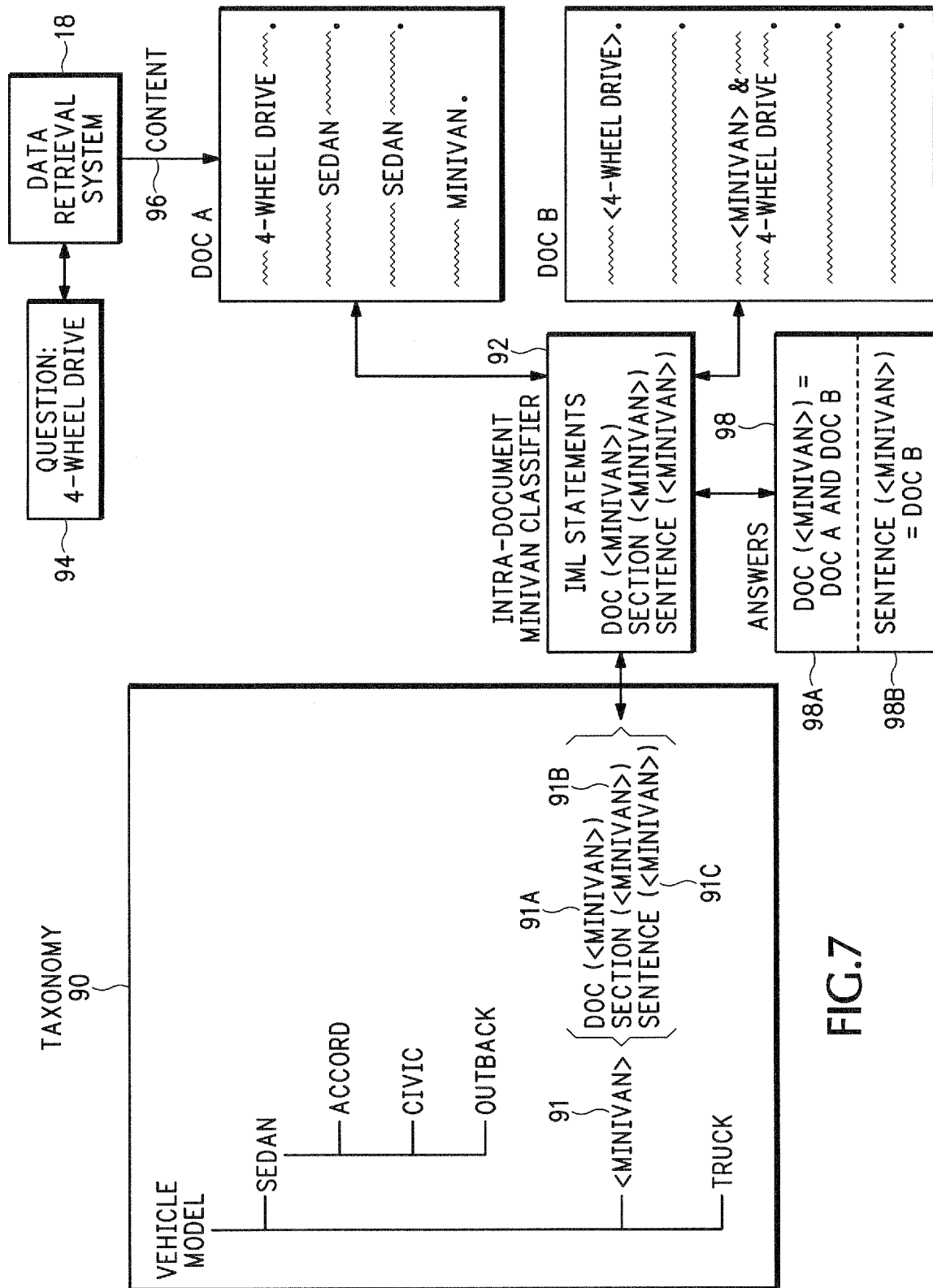


FIG. 7

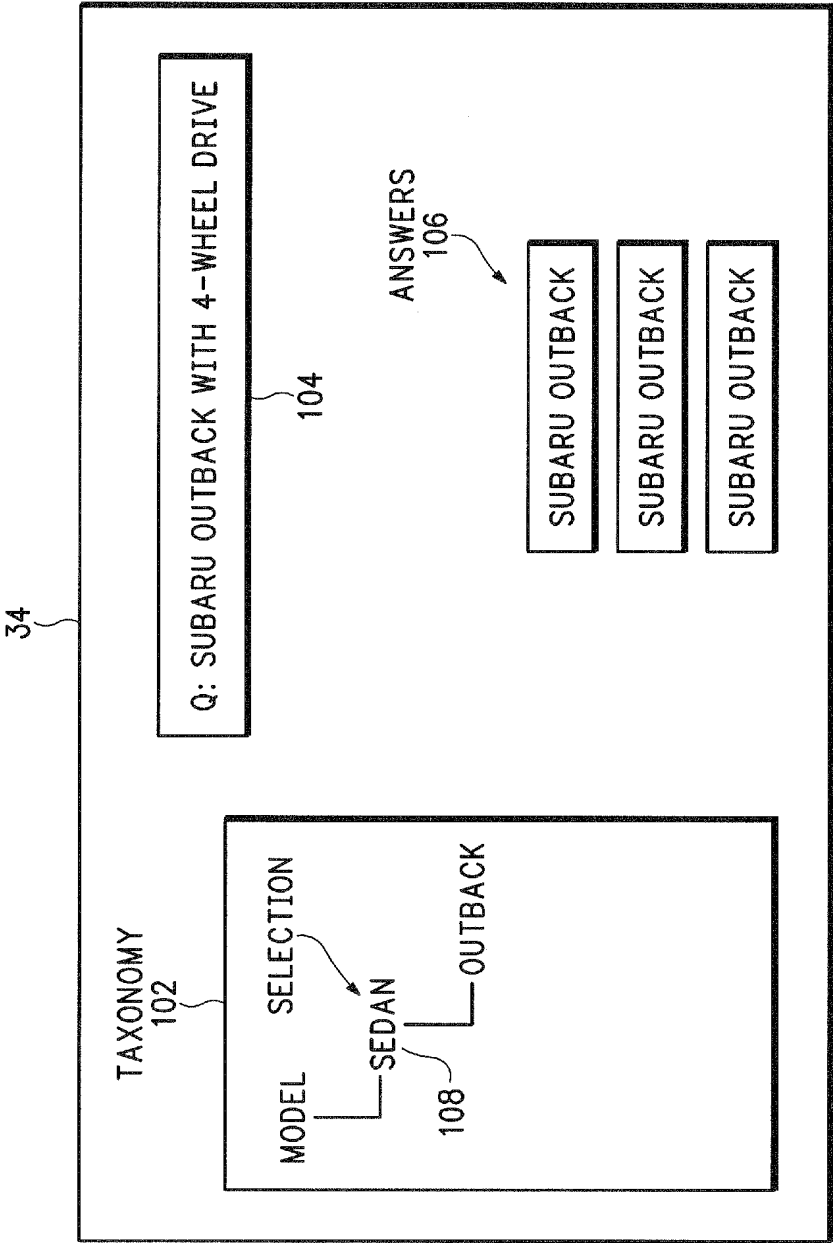


FIG.8

9/17

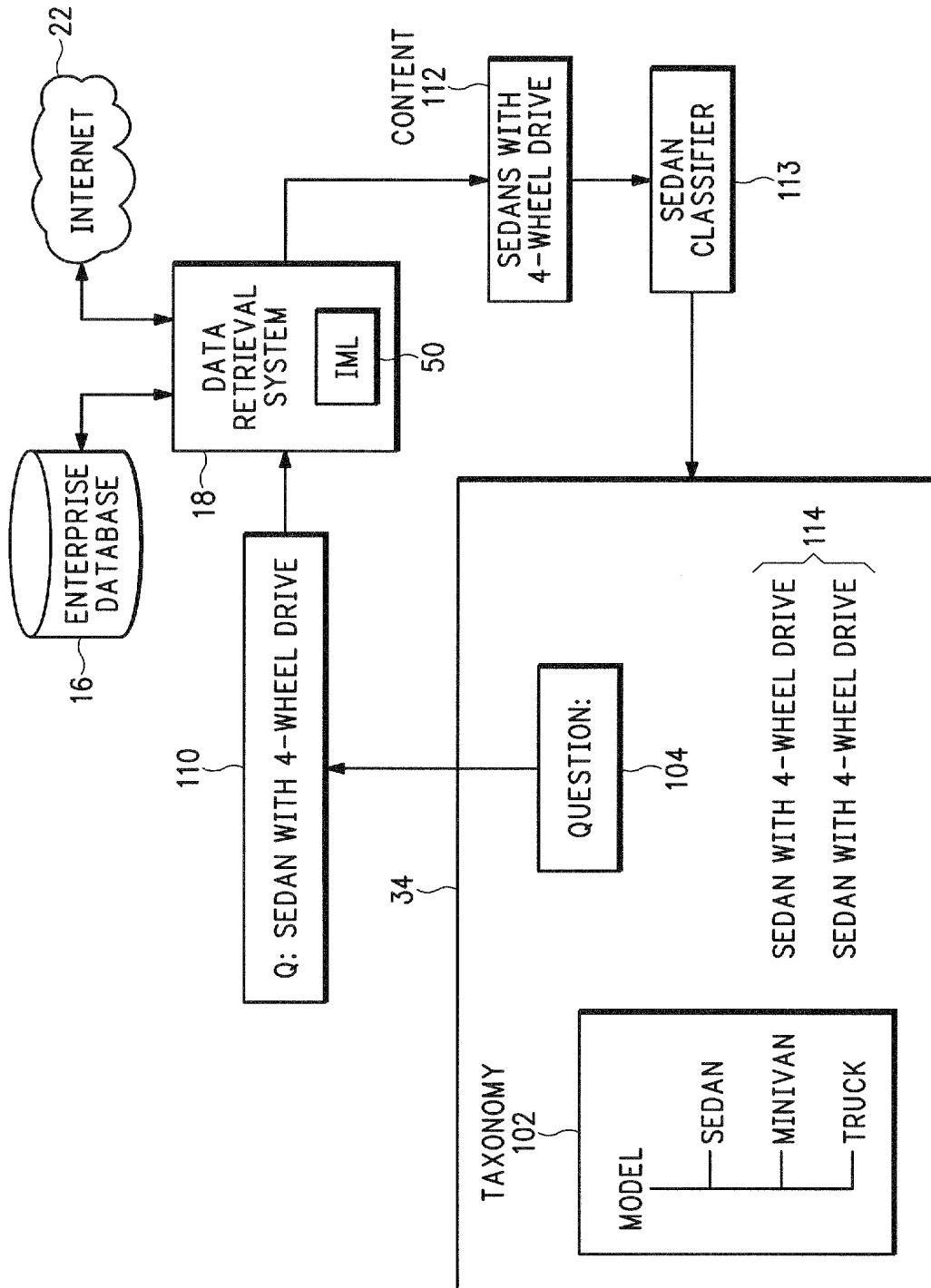
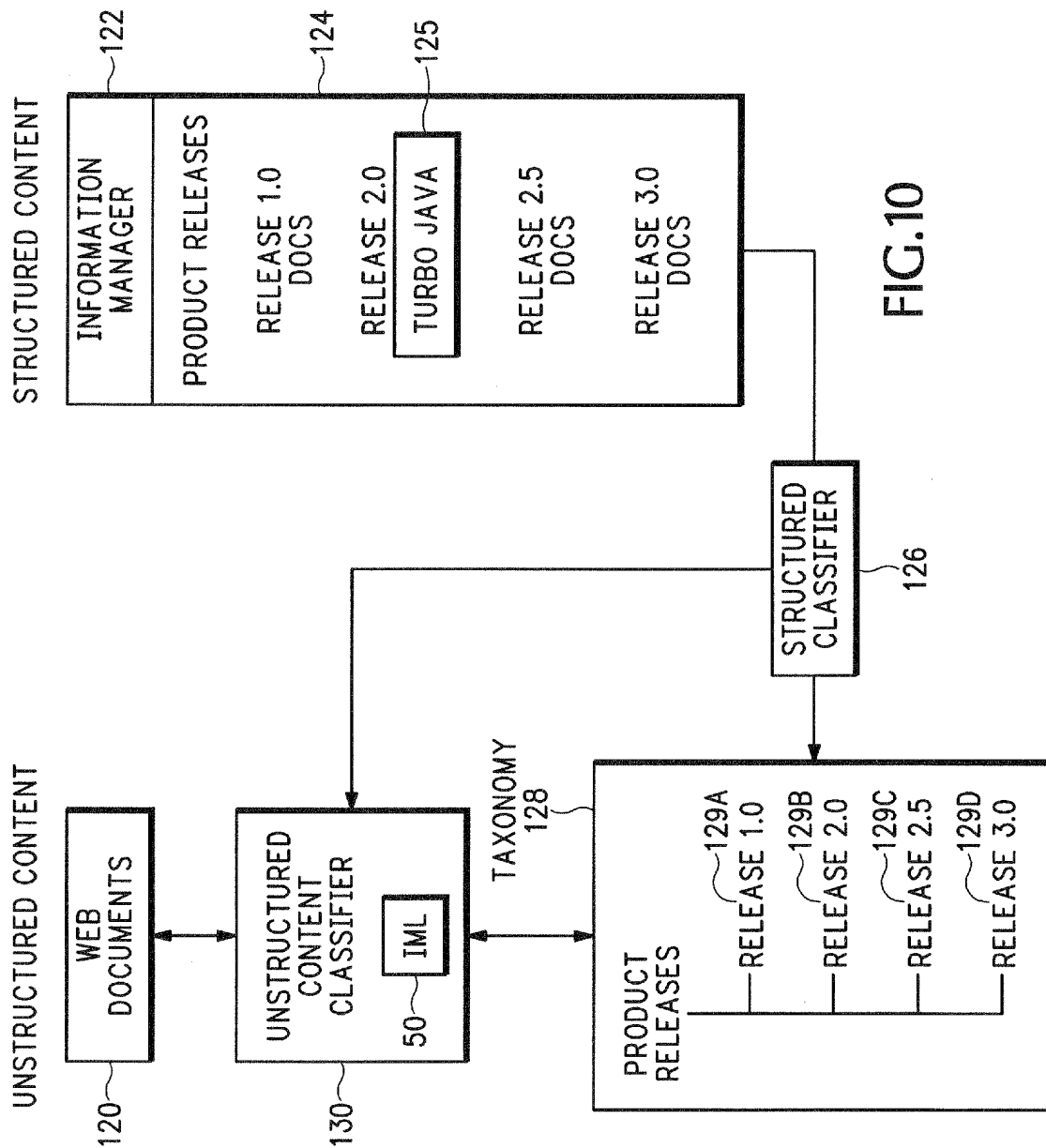
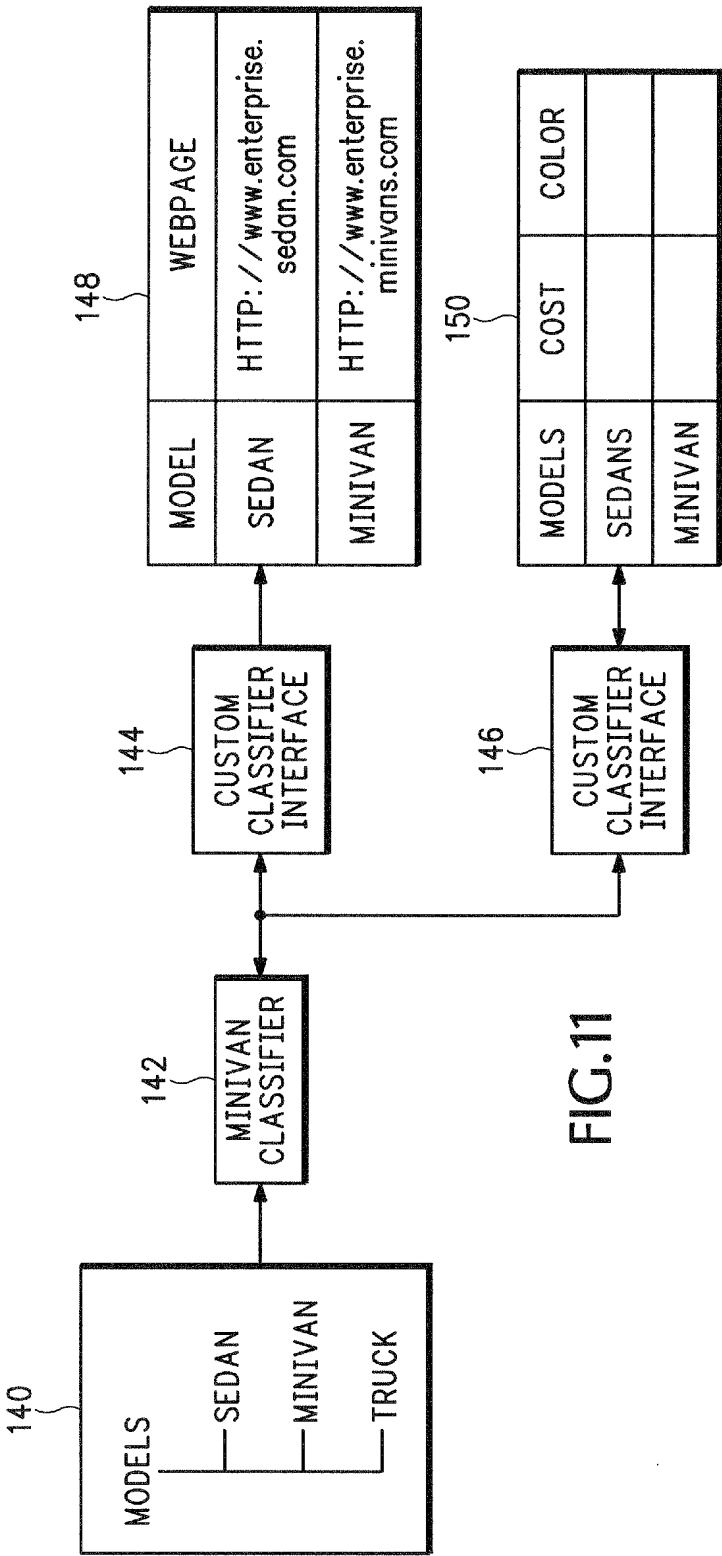
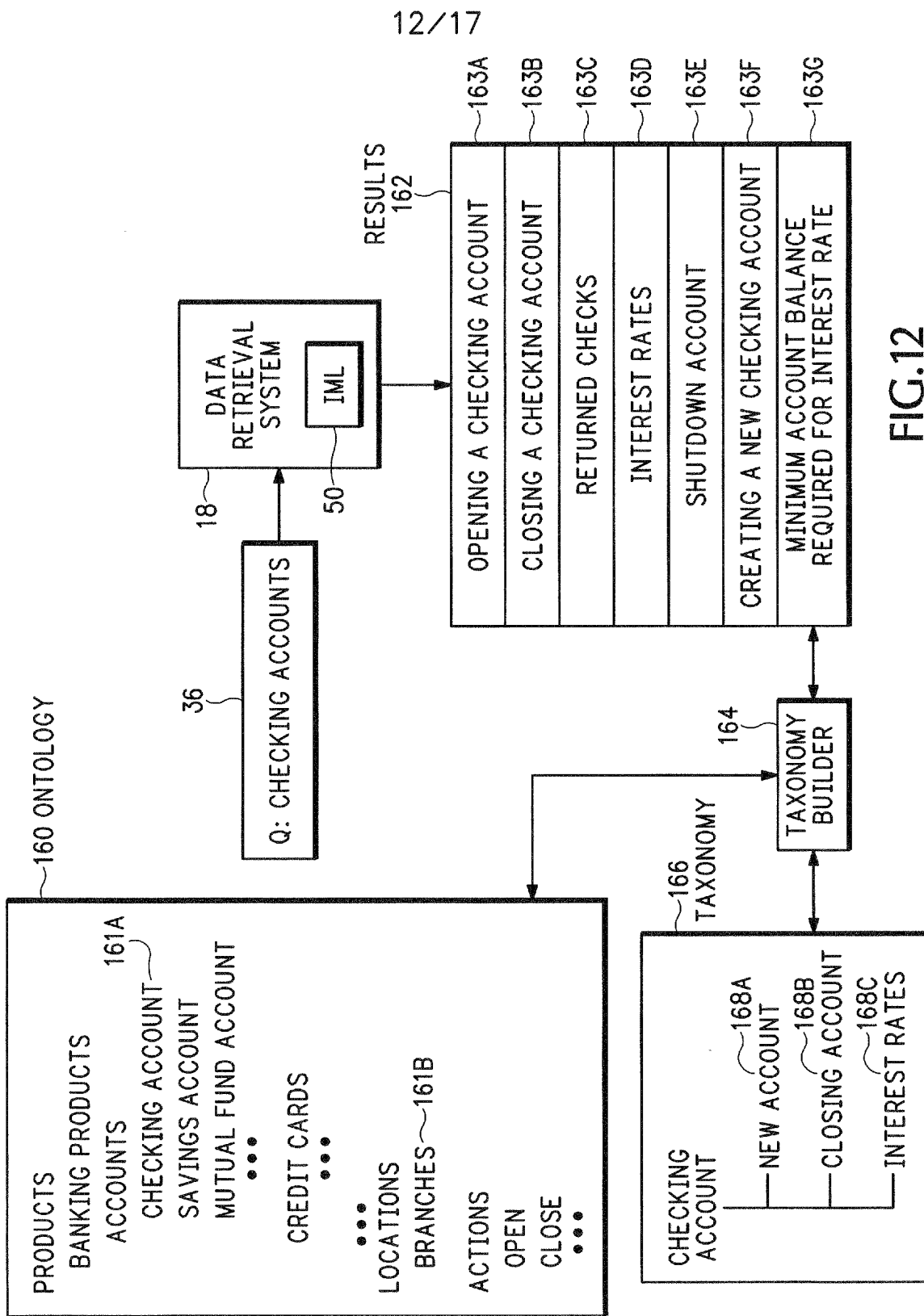


FIG.9

10/17







13/17

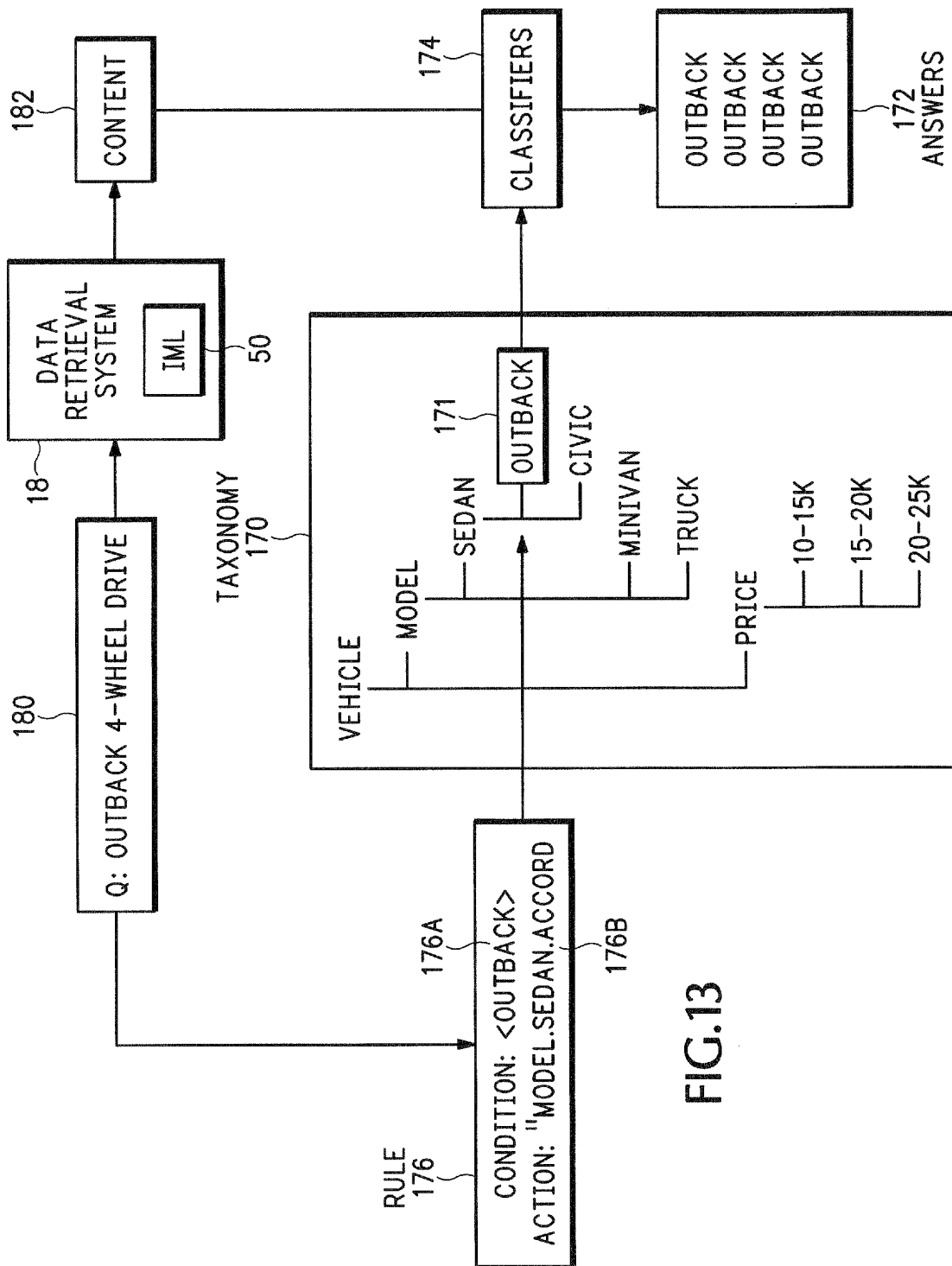
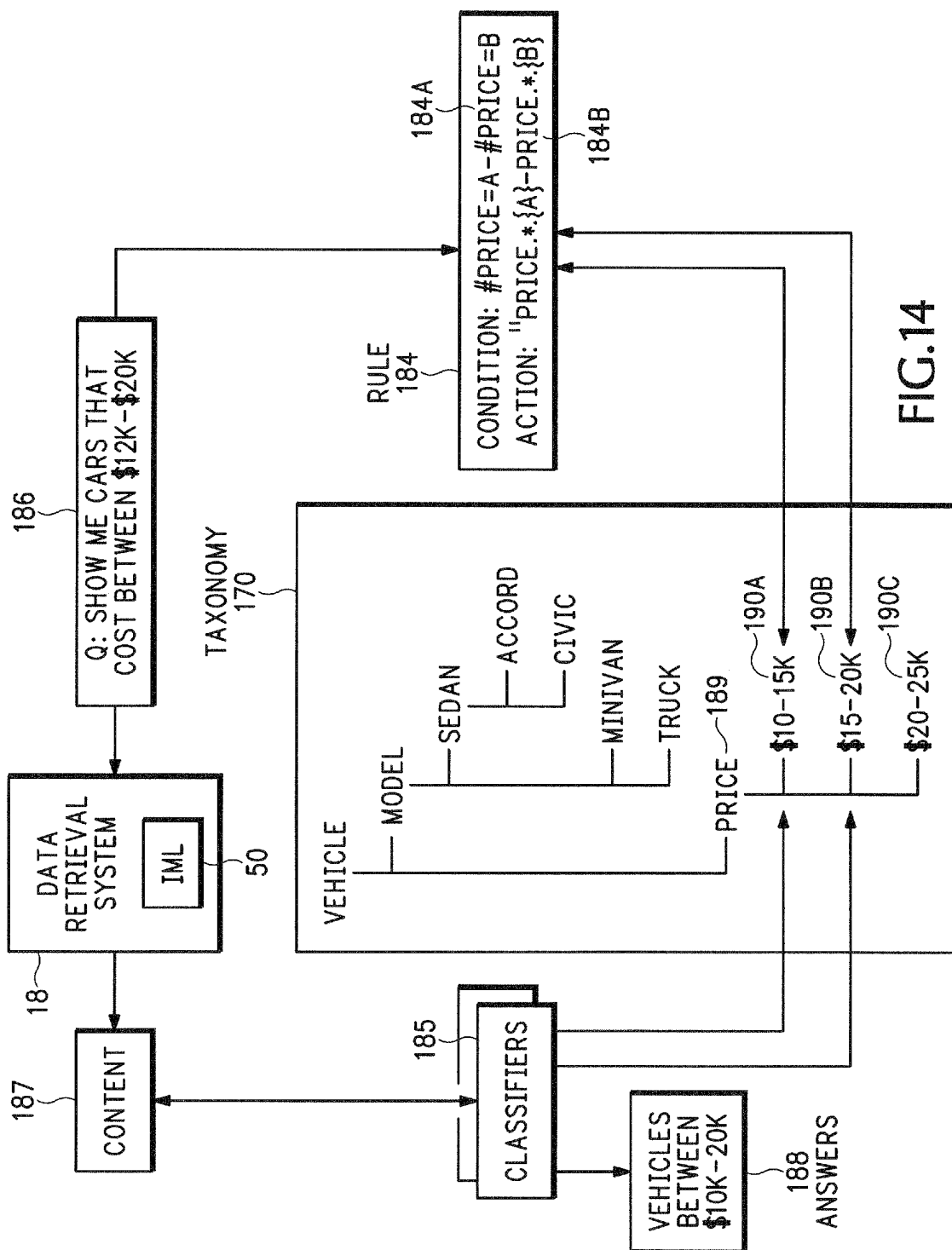


FIG.13

14/17



15/17

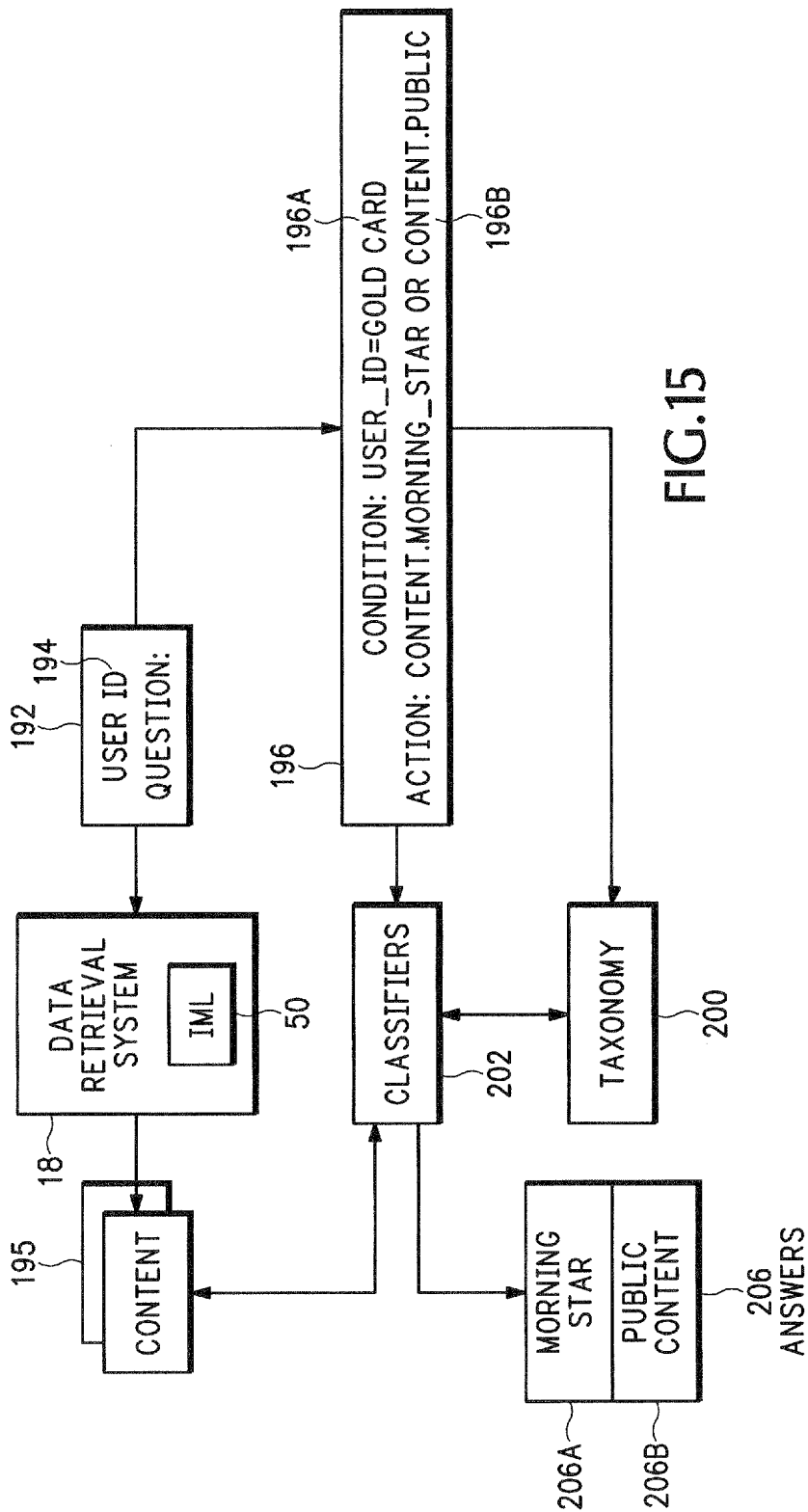


FIG.15

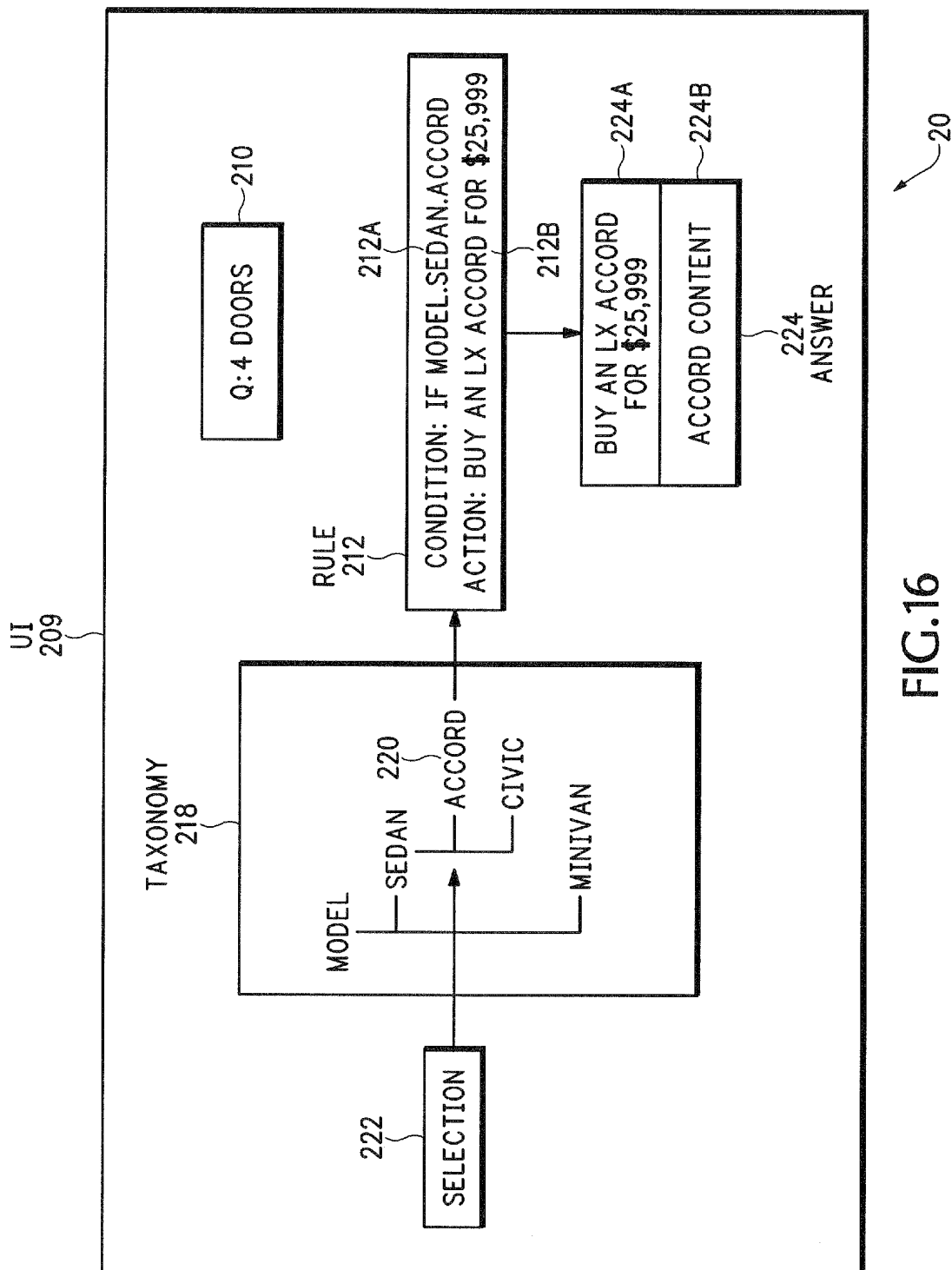


FIG.16

17/17

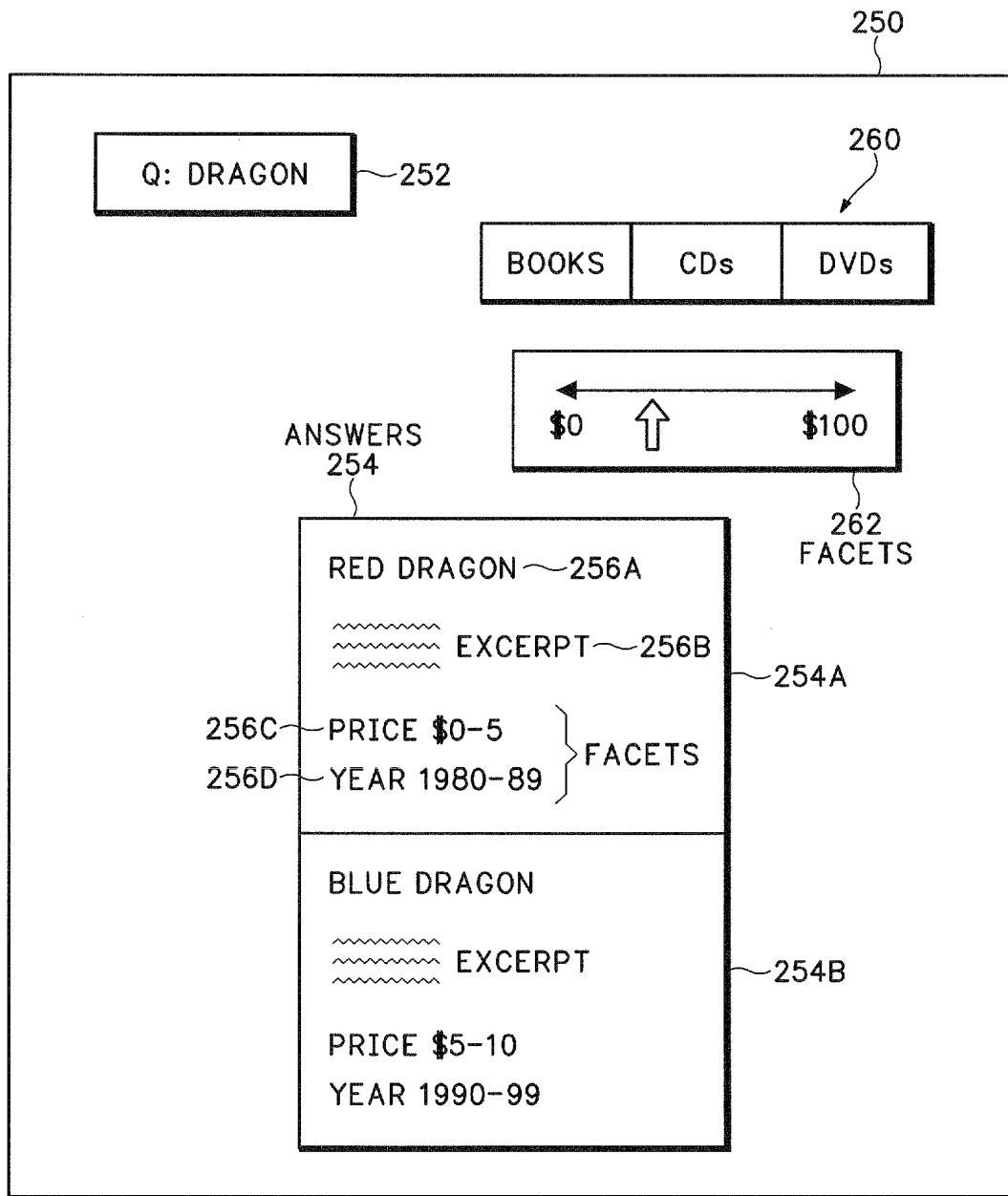


FIG.17